MEASURING MOTOR ABILITY

A SCALE OF MOTOR ABILITY TESTS

GV436 .B7
Measuring motor ability : a scale o
Gutman Library ANR4387
3 2044 028 700 292

DAVID KINGSLEY BRACE

HARVARD UNIVERSITY

LIBRARY OF THE

GRADUATE SCHOOL
OF EDUCATION

Digitized by the Internet Archive
in 2023 with funding from
No Sponsor

WITHDRAWAL

https://archive.org/details/measuringmotorab0000davi

MEASURING MOTOR ABILITY

MEASURING MOTOR ABILITY

A SCALE OF MOTOR ABILITY TESTS

BY

DAVID KINGSLEY BRACE, PH.D

Professor of Physical Education
University of Texas, Austin, Texas.

NEW YORK
A. S. BARNES AND COMPANY
1927

HARVARD UNIVERSITY
GRADUATE SCHOOL OF EDUCATION
MONROE C. GUTMAN LIBRARY

GV436
.B7

JUNE 22, 1938

COPYRIGHT, 1927, BY
A. S. BARNES AND COMPANY

PRINTED IN THE UNITED STATES OF AMERICA

ACKNOWLEDGMENT

THE writer takes pleasure in acknowledgment of his indebtedness to those who have assisted in making possible this study. Sincere thanks are given to Teachers College for its inspiration for research, and to the many friends whose encouragement and advice have been so valuable and deeply appreciated.

The writer is especially grateful to Dr. J. F. Williams for his continued support and guidance; and to the other members of his advisory committee: Professors Henry A. Ruger, Arthur I. Gates and Frederick G. Bonser for their criticisms and suggestions; and to Professor William H. McCall for his invaluable aid and advice.

Especial thanks are due to Mr. Carl Stockdale, Director of Physical Education, Baldwin, Long Island; Mr. Colba F. Gucker, Director of Physical Education, Lincoln School, New York City; Miss Mary Shafer, Teachers College; Dr. Clara M. Greenough, Skidmore College, Saratoga Springs, New York; and Miss Catherine Avery, New Jersey College for Women, New Brunswick, New Jersey, for their assistance by giving the tests in their schools.

The writer desires to gratefully acknowledge the constant help of his wife, Dorothy Walton Brace, whose skillful and painstaking assistance has made the completion of this study possible.

D. K. B.

TABLE OF CONTENTS

CHAPTER		PAGE
	PREFACE	1
I.	THE TESTS AND THE SUBJECTS	1
	THE TESTS USED	2
	THE NATURE OF THE TESTS	5
	SUBJECTS	7
	SUMMARY	8
II.	ADMINISTERING THE TESTS	10
III.	THE VALIDITY OF THE TESTS AS MEASURES OF MOTOR ABILITY	13
	SELECTION OF CRITERIA FOR DETERMINING THE VALIDITY OF THE TESTS AS MEASURES OF MOTOR ABILITY	16
	JUDGMENT RATINGS AS A CRITERION OF MOTOR ABILITY	17
	THE VALIDITY OF THE MOTOR ABILITY TESTS SHOWN BY CORRELATIONS WITH ACHIEVEMENT IN ATHLETIC EVENTS	21
	MOTOR ABILITY SCORES OF TEAM MEMBERS	24
	SUMMARY	25
IV.	THE DETERMINATION OF RELIABILITY AND OBJECTIVITY	27
	CHANGES IN EQUIPMENT NOT A CAUSE OF UNRELIABILITY	27
	INSTRUCTIONS FOR ADMINISTERING THE TESTS ARE CLEAR AND DEFINITE	28
	COMPARISON OF FAILURES ON FIRST TRIAL WITH FAILURES ON A RETRIAL	29
	THE RELIABILITY OF PUPILS' SCORES	33

CHAPTER PAGE

Correlation: Self-Correlation Between Two Approximately Equivalent Forms of the Motor Ability Test 35
The Mean Net Difference Between Forms A and B 36
The Mean Net Difference Between Two Sets of Motor Ability Tests Over an Intervening Year . . . 36
Correlation: Coefficient of Correlation for Scores on Motor Ability Tests One Year with Scores on the Same Tests and Same Students One Year Later 38
Correlation: Failures on First Trial with Failures on Second Trial . 40
Correlation: First Trial with Second Trial Scores Made by College Women 41
Objectivity 42
Summary 45

V. Scaling the Tests 48
Determination of the Optimum Scoring Key 48
Weighting Tests in Accordance with Their Difficulty 49
Weighting Tests on the Basis of Mean Criterion Scores 49
Weighting the Tests on the Basis of Difference Between Mean Criterion Scores for Successes and the Mean Criterion Scores for Failures 52
Weighting Scores on the Basis of the Validity of Each Test 54
Selection of Tests to be Included in the Final Scale 59

CHAPTER PAGE

Construction of Equivalent Forms . 63
Selection and Application of a Scaling Technique 66
Summary 70

VI. Some Applications of the Scale of Motor Ability Tests 73
Correlation: Scores on Certain Athletic Game Skills and Athletic Events with the Motor Ability Tests 73
Achievement Tests in Basketball Skills 74
Achievement Tests in Indoor Baseball 78
Correlation: Achievement Tests in Indoor Soccer-Football with the Motor Ability Tests 81
Correlation: Scores on the Motor Ability Tests with Scores on Indoor Track and Field Events . . . 85
Correlation: Motor Ability Tests with Age 86
Correlation: Motor Ability Tests with Weight 87
Correlation: Motor Ability with Intelligence 88
Suggested Applications 89

VII. General Summary 93

VIII. Test Booklet—Brace Scale of Motor Ability Tests 100
Illustrations of the Tests with Descriptions and Instructions on Scoring 105
Scoring Blank 125

Appendix I: Age Distributions . . . 129
Appendix II: Charts 131
Bibliography 133
Index 135

INDEX TO TABLES

NO.		PAGE
I.	TYPES OF ACTIVITIES SAMPLED BY THE TESTS	6
II.	JUDGMENT CRITERIA OF GENERAL MOTOR ABILITY RANKED BY PHYSICAL EDUCATORS AS TO VALUE	14
III.	RELATIVE VALUE OF 10 MOTOR ACTIVITIES IN WHICH SKILL CAN BE TAKEN AS A CRITERION OF GENERAL MOTOR ABILITY	15
IV.	SHOWING HOW JUDGMENT RATINGS OF MOTOR ABILITY WERE TO BE GIVEN .	18
V.	CORRELATION COEFFICIENTS FOR SCORES ON THE MOTOR ABILITY TESTS AND JUDGMENT RATINGS FOR BOYS AND GIRLS IN THE BALDWIN GROUP . .	20
VI.	FAILURES ON SUCCESSIVE TRIALS AND PER CENT OF CONSTANCY	30
VII.	DETERMINING APPROXIMATE RELIABILITY BY THE MEAN NET DIFFERENCE METHOD APPLIED TO SUCCESSIVE TRIALS	32
VIII.	PER CENT OF AGREEMENT IN GROUP SCORING BY SIXTH GRADE BOYS . . .	34
IX.	DETERMINING THE UNRELIABILITY OR MEAN NET DIFFERENCE BETWEEN FORM A AND FORM B	37
X.	THE MEAN NET DIFFERENCE BETWEEN TWO SETS OF MOTOR ABILITY TEST SCORES OVER AN INTERVENING YEAR	38
XI.	CORRELATIONS INDICATING OBJECTIVITY AND RELIABILITY IN GROUP SCORING .	44

NO.		PAGE
XII.	Tests Weighted in Accordance with Difficulty	50
XIII.	Mean Criterion Scores for Tests Passed	51
XIV.	Mean Criterion Scores for Tests Passed and Tests Failed	53
XV.	Illustrating the Determination of the Sum of Criterion Scores for a Single Test Item	56
XVI.	Illustrating the Determination of the Validity of a Single Test Item	57
XVII.	Data Relating to the Determination of the Validity of Each Test	58
XVIII.	Coefficients of Correlation Depending upon the Number of Test Elements Used	61
XIX.	Data Used in Eliminating Certain Tests for the Final Scale	62
XX.	Balance Tests in Forms C and D	64
XXI.	Computation of a T Scale for 11-12-13 Year Olds	68
XXII.	Extending the Range of the T Scale	70
XXIII.	T Scales for Six Basketball Achievement Tests	76
XXIV.	T Scales for Baseball Achievement Tests	80
XXV.	T Scales for Soccer Achievement Tests	84
XXVI.	Approximate Meaning of Scale Scores	104
XXVII.	Scoring Blank for Brace Scale of Motor Ability Tests	125
XXVIII.	Tables for Converting Test Scores into Scale Scores	126
XXIX.	Scoring Table for College Women	127
XXX.	Age Distributions, Boys and Girls, Ages 8 to 18	129
XXXI.	Age Distributions for College Women, Ages 13 to 48	130

INDEX TO CHARTS

NO.		PAGE
I.	SCATTERGRAM: SCORES ON MOTOR ABILITY TESTS WITH JUDGMENT RATINGS . . .	131
II.	SCATTERGRAM: FIRST TRIAL FAILURES WITH FAILURES ON SECOND TRIALS OF THE SAME TESTS	132

PREFACE

Tests and measurements have long been associated with various phases of what is now known as physical education.

The present educational measurement movement finds interested support among physical educators. While many tests are being used there is a lack of standardized educational tests in physical education Certain tests of physical capacity have been standardized [1] and show high correlations with some athletic abilities.[2] So-called physical efficiency tests [3] prove on examination to be largely unscientific and to be, in reality, more probably achievement tests. The Sargent "Physical Test of a Man," [4] and the improvement on this test by Schwegler and Engelhardt,[5] are interesting studies. The "Decathlon Tests," developed by Hetherington,[6] provide percentile scales which doubtless measure, in part, motor ability, although in the main they serve the purpose of achievement tests.

A modern program of physical education should provide for motor achievement, using the term motor to designate those muscular reactions resulting in manipulating the body. Older programs were concerned with structure rather than function, in developing large muscles, in increasing strength, and, in general, in increasing physical capacities. The more modern workers in this field are interested first of all in activ-

[1] Whipple, G. M.—A Manual of Mental and Physical Tests, Part I, pp. 92ff.

[2] Rogers, F. R.—Physical Capacity Tests in the Administration of Physical Education.

[3] Proceedings of the Society of Directors of Physical Education in Colleges, 1921.

[4] Sargent, D. A.—School and Society, p. 128, January 29, 1921,

[5] A Test of Physical Efficiency, American Physical Education Review, November, 1924.

[6] Published by the American Physical Education Association Press.

ities which have meaning and are worthwhile for those engaging in them.[7]

The measurement movement in physical education has, however, been greatly handicapped at the outset by the absence of any standardized tests of fundamental motor ability which individuals must use in learning and performing physical activities. Such motor ability may be thought of as corresponding on the physical side to mental ability, on the intelligence side of human activity.

The lack of any scientific measurement of motor ability and the hope of helping to supply this need has led the author to attack the problem of developing the Scale of Motor Ability Tests presented in this book.

The statistical procedures used in developing and validating the Scale of Motor Ability Tests are necessarily included in a book of this nature. Certain achievement tests in athletic game skills are included for purposes of clarifying some of the applications of the motor ability tests.

It is hoped that this Scale of Motor Ability Tests may assist physical educators in the classification of their pupils, in furnishing a basis upon which to evaluate achievement, and in gaining a better understanding of the motor side of human reactions.

Illustrations of the tests and complete information for administering and scoring the scale tests are collected in Chapter VIII in order to be of ready reference to readers who may wish to give the Brace Scale of Motor Ability Tests.

DAVID KINGSLEY BRACE.

January, 1927.

[7] Williams, J. F.—Physical Education.

MEASURING MOTOR ABILITY

CHAPTER I

THE TESTS AND THE SUBJECTS

THE test elements of which the scale is comprised have been selected by the empirical method assisted by the following guiding principles and criteria:

(1) The tests should primarily test motor ability.

(2) The tests should test native ability rather than acquired ability.

(3) The tests should involve a general functioning of the body musculature.

(4) The tests should sample a variety of types of reactions.

(5) The tests should be easy of administration and simple to score.

With these criteria in mind, a collection of possible test elements was started and a group of more than seventy tests was obtained.

In accordance with the criteria above, all tests were eliminated from consideration which measured acquired skills such as throwing basketball goals, high jumping, etc. Tests which involved understanding of complicated instructions were dropped. Tests which involved a special and restricted muscular reaction, such as tapping, tracing, and spearing, were eliminated. Tests which were difficult to score or required the use of any equipment whatever were discarded.

The tests which were finally selected require no equipment other than pencils and scoring blanks. They are scored simply "success" or "failure." The total test can be conducted as an individual test or as a group test. As a group test, children may either score themselves or score each other. From ages above ten years, the total test can be easily given to a group of forty to one hundred or more pupils by one examiner in a

forty-minute period. However, with the younger ages, smaller groups are preferable.

Below the fifth grade the test should be given as an individual test. However, the average teacher of physical education should be able to score four children simultaneously. A still larger number can be scored at one time if they perform each test one at a time. If this latter plan is used, care must be taken to rotate the order in which pupils try the test so as to equalize any benefit that may accrue from watching others.

The Tests Used

The test elements selected and treated in this study are briefly described below.[1]

BRIEF DESCRIPTION OF THE TESTS[2]

Form A[3]

1. Stand still, with feet together, close both eyes. Hold the position for ten seconds without moving. Start when I say "go," and stop when I say "stop."

2. Holding the left foot in the right hand behind the right leg, hop around on one spot in a circle three times without losing balance.

3. Clasp hands on the head, step forward with the left foot and kneel onto the right knee and return to a stand, without moving the feet from the first position.

4. Stand, kick the right foot up so that the toe comes at least level with the shoulders, without falling over.

5. Jump into the air and clap the feet together once and land with feet at least two inches apart.

6. Jump into the air and make a half turn to the right,

---

[1] Subsequent experimentation caused the dropping of certain tests and the arrangement of those remaining into two approximately equivalent forms, as explained in Chapter V and described in Chapter VIII, p. 105 ff.

[2] For full instructions and illustrations, see Chapter VIII.

3. The tests in Form A are of approximately equal difficulty with those in Form B. Tests in each Form are arranged in order of difficulty.

landing on the same spot (but facing about) without losing the balance, *i.e.*, without moving the feet after they first hit the floor.

7. Lie flat on the back on the floor, fold the arms on the chest, raise the trunk to a sitting position without unfolding the arms.

8. Fold the arms behind the back, kneel onto both knees, and get up without losing balance or moving the feet about.

9. Stand on the left foot, close the eyes. Hold the position without moving for five seconds. Start when I say "Go," and stop when I say "Stop."

10. Stand with both feet together, bend down, extend both arms down between the knees, around behind the ankles, and hold the fingers together in front of the ankles without losing balance. Hold this position for five counts.

11. Take full squat position with arms together in front of the body, fingers touching the floor. Spring up onto both heels, with toes up, swinging both arms at side horizontal, legs straight, and feet about 18 inches apart. Repeat this three times rhythmically.

12. Jump into the air and make a full turn left, landing on the same spot and not losing balance, *i.e.*, moving feet after they first hit the floor.

13. Jump into the air and clap the feet together twice and land with feet at least two inches apart.

14. Stand on the left foot with the right foot extended forward off the floor. Squat on the heel of the left foot without touching the right foot or hands to the floor, and stand without losing the balance.

15. Hold either foot in the opposing hand and jump through the loop thus made.

Form B

16. Stand on the left foot, with eyes open, for five seconds without moving. Start when I say "Go" and stop when I say "Stop."

17. Walk in a straight line, placing the heel of one foot in front of and against the toe of the other foot. Place each foot on the floor five times, without losing the balance. Begin with the left foot, *i.e.,* step ten times.

18. Stand with hands on hips, go to a full squat position with heels off the floor, keeping trunk straight and perpendicular. Return to a standing position slowly without losing balance or bending the trunk forward, or moving the hands.

19. Jump into the air and slap both heels with both hands behind the back.

20. Walk backward in a straight line for ten steps, placing the toe of one shoe against the heel of the other. Look at the feet, but do not lose the balance or fail to walk in a straight line.

21. Touch the tips of the fingers to the floor without bending the knees.

22. Swing the arms and jump up in the air with a *half* turn left, coming down on the same spot and not losing the balance.

23. Stand on the left foot, place both hands on the floor in front of the left foot, raise the right leg and extend it behind, touch the head to the floor, and regain the standing position without losing balance.

24. Take a front leaning rest position, *i.e.,* rest on the floor supporting the body by both hands and feet with arms and body straight. Bend the arms, touching the chest to the floor, and push up again to straight arms *three* times in succession. *Do not allow knees or waist to touch the floor.*

25. Swing the arms and jump up in the air, making a full turn to the right, landing on the same spot and not losing the balance, *i.e.,* without moving the feet after they first hit the floor.

26. Kneel onto both knees, extend the toes out flat behind, swing the arms and jump to the feet without rocking back on the toes, or losing the balance.

27. Fold arms on chest, sit down cross-legged and get up

without unfolding the arms or having to move the feet about to regain the balance.

28. Stand on the left foot, hold the right foot against the left knee, place both hands on hips, shut both eyes, hold the position for ten seconds without shifting the left foot on the floor.

29. Stand on the right foot, grasp the left foot behind the right knee, touch the left knee to the floor, and stand without touching any other part of the body to the floor, or losing the balance.

30. Take a squat leaning position with elbows out, *i.e.*, squat on the heels with hands on the floor in front of the feet, place the knees well over the elbows and rock forward onto the hands, raising the feet off the floor. Hold the position for five seconds.

The Nature of the Tests

From the description of the tests it will be seen that they are of the nature of stunts. They call for coördinations, more or less general, a factor which is increased by their variety.

It is the opinion[4] of the author that the tests represent various types of activities as shown in Table I, page 6.

The classification of Table I, page 6, is however, a rough one and is used merely to assist in understanding the tests. Test Number 10 may be taken as an illustration. This test is called a test of control. It calls for bending forward and placing both hands on the floor, raising the right leg and touching the head to the floor. It is partly a test of balance control, partly a test of strength of arms and partly a test of abdominal and back muscles. Since these elements are perhaps present in about equal amounts, the term control has been used.

For purposes of this study it is not necessary to analyze

[4] No objective proof is presented in support of this opinion.

TABLE I

TYPES OF ACTIVITIES SAMPLED BY THE TESTS *

Types of Activity	Test Number	Frequency by Tests
Agility	5, 11, 13, 15, 19, 26	6
Balance	1, 9, 16, 17, 20, 28	6
Control	3, 8, 10, 18, 23	5
Flexibility	4, 21	2
Agility and Balance	2, 6, 12, 22, 25	5
Strength	7, 24	2
Strength and Control	14, 27, 29, 30	4

* Based upon the author's judgment.

closely the types of activity involved in the tests. The rough analysis suggested in Table I will suffice to show that the tests do sample a wide range of activity.

A wide variety of muscular reactions are called for. The abdominal muscles are fundamental in many of the activities. That these muscles play a major importance in most active bodily movements is well recognized. Four of the agility tests, 5, 11, 13, 15, 19, 26, require jumping accompanied by active movements of the arms or legs, or in the case of tests 11 and 26, both arms and legs. In test 5 the jump is from the knees practically unassisted by the feet. The lift must come from the trunk muscles and arm swing.

Tests 4 and 21 sample two varieties of stretching. In tests 14, 27 and 29 strong leg muscles are required, although it is again clear that a general reaction by the body as a whole is demanded. Test 7 requires a sit-up from a prone lying position and tests the abdominal muscles primarily. Tests 30 and 24 tax principally the strength of the arms. The tests

of balance call for reactions, muscular and otherwise, involved in standing.

Subjects

The Motor Ability Test can be used for either boys or girls, men or women. The elements are best performed by boys with coats off and in tennis shoes or other light rubber soled shoes. Older girls or women should be in gymnasium costume and shoes to secure best results.

The Test has been given satisfactorily to children of 8 years of age and to women of forty-eight years of age. It will test ability anywhere along this range.

A total of 1,555 students have been given the tests. Of this number 1,309 have scores on the Scale of Motor Ability Tests.[5] The scores of 246 pupils could not be considered in the age norms, either because they had failed to give their age or had not taken all tests.

Subjects whose scores have been used in this study are distributed by groups as follows:

(1) Public Schools of Baldwin, Long Island
  (a) Ages 10 to 18
    Boys—231; girls—220; total..... 451
  (b) Ages 12 to 15
    Boys—67; girls—50; total..... 117[6]
(2) Horace Mann School, New York City
  (a) Ages 9 to 13
    Boys—70; total..... 70
  (b) Ages 9 to 13
    Boys—61; total..... 61[7]
(3) Lincoln School, New York City
  (a) Junior and Senior High School
    Ages 8 to 18
    Boys—188; total..... 188[8]

[5] Age norms for these pupils appear in Appendix I, p. 129.

[6] These individuals appear twice, having been tested twice with an interval of one year between testings.

[7] Not included in Table XXX, p. 129, Age Distributions.

(b) Elementary School
Ages 8 to 13
Boys—125; total ..... 125[8]

(4) Teachers College, New York City
(a) Women in Extension Classes
Ages 17 to 48
Women—69; total..... 69
(b) Junior and Senior Women
Majoring in Physical Education
Ages 17 to 35
Women—255; total..... 255
(c) Men in Physical Education
Ages 24 to 32
Men—7; total ..... 7[9]

(5) Skidmore College, Saratoga Springs, N. Y.
College students, four years
Ages 13 to 19
Women—145; total..... 145

(6) New Jersey College for Women, New Brunswick, N. J.
Four classes
Ages 16 to 23
Women—67; total..... 67

GRAND TOTAL.....1555

The number of tests passed by students of each age is shown in Table XXX, page 129.

Summary

(1) The test elements selected for this investigation were chosen because they appeared to conform with certain criteria. They are intended to test motor ability of a general sort and native rather than acquired.

[8] One hundred and fifty-seven of these two groups are not included in Table XXXI, p. 130, Age Distributions.

[9] Not included in Table XXX, p. 129, Age Distributions.

(2) The Scale of Motor Ability Tests is intended to be given as an individual test or as a group test. As a group test children may score themselves or be scored by each other.

(3) No equipment is required except a watch, pencils and score blanks.

(4) The tests sample a variety of elements such as agility, balance, control, flexibility, and strength.

(5) Data used in this study was obtained from examination of 1,555 students from the fourth grade through college and from ages 8 to 48 years.

Age norms given in a subsequent chapter are based upon scores of 1,298 students.

CHAPTER II

ADMINISTERING THE TESTS

DETAILED instructions for administering the Motor Ability Tests are given in the Test Booklet in Chapter VIII. A brief explanation of the procedure in giving the tests is presented here so that the reader may have a better understanding of following chapters.

The thirty-three test elements, or tests, selected as explained in the preceding chapter were carefully described and put in mimeographed form. A slight attempt was made to arrange the tests in an approximate order of difficulty. They were also arranged so that continued exercise of the same muscle groups would be lessened.

Instructions for giving the tests were carefully prepared, tried out and mimeographed.[1]

In brief, the procedure in giving the tests as a group test is to assemble the class in a gymnasium, room, hall or play yard, line up the class, and by counting off by two's, or by marching, or by any other method, arrange the pupils in as many double lines or columns of two's as the space permits. These double lines then face together. Scoring blanks and pencils are passed out from the ends of the lines.

Children are then instructed to sit on the floor and fill in the spaces at the top of their score sheets, giving their names, school, class, grade, age, weight, sex, and the date.

Papers are then exchanged, so that two children face each other, each holding the scoring blank of the other. Alternate lines are numbered one and two.

While the children are seated, the nature of the test is explained to them. In general,[2] this explanation is as follows:

[1] See Chapter VIII, p. 100.

[2] See Chapter VIII, p. 100 for specific directions.

You are going to be given a test consisting of a number of tests or stunts. Some are easy and some are difficult. Probably no one will be able to do them all, but all of you will be able to do many of them.

You are to score each other in this test. Each stunt will be explained and demonstrated. When you try the test your scorer will give you an X if you do the stunt, and an O if you fail to do it exactly as explained. Each test is to be scored either "success" or "failure." If the test is done correctly, mark an "X" in the space by the number of the test. If the test is failed, mark an "O" [3] in the space by the number of that test. Each test must be marked either "X" or "O," success or failure.

Is there any question about how the tests are to be scored? For example, the first test is to stand on both feet, close both eyes and hold the position for 10 seconds without losing the balance. Start when I say "go" and stop when I say "stop". Suppose I do it like this and move my feet like that.[4] That would count as failing the test and I should be scored zero.

Lines Number Two are then requested to lay pencils and papers on the floor and to stand. The first test is explained and demonstrated.[5] All those standing are then instructed to do the test on the commands "Ready—Go." The tester then says:

Each scorer mark the person you are scoring. If he (or she) did not move the feet about on the floor, mark a cross for success. If he (or she) moved his feet, mark a zero for failure. Is there any question about how to score the test?

This process is then repeated for the first five tests, after which Lines Two sit down and Lines One stand. The first five tests are then given to those pupils in Lines One, who remain standing and take the next five tests also. This method

[3] Demonstrate this mark by pantomime.

[4] Demonstrate the movement.

[5] It is not essential that the tester be able to perform each test. An attempt at performance will serve to show what is required, or one pupil may be called out to demonstrate. However, in case of doubt, if the instructions are simplified, the children will be able to learn what is wanted without any demonstration.

of alternating between the two halves of the class is continued until all have taken the test.[6]

When all have taken the tests, score sheets and pencils are passed to the front of the lines and collected by two pupils assigned to this duty.

No practice trials are allowed, and it is not necessary to give a second trial, though the scoring blank provides spaces for second trials if the tester wishes to give them.[7]

A pupil's motor ability score consists of that Scale Score which corresponds to the number of tests passed. Scale Scores are to be obtained from Table XXVIII, page 126, Chapter VIII.

The reader should note carefully that Scale Scores are given only for records made on those tests in Forms M and N which comprise the Brace Scale of Motor Ability Tests.[8]

---

[6] In the final method, Chapter VIII, p. 102, for administering the test, Form M can be given to one set of lines and Form N given next to the other set of lines. This order is then reversed so that all are given the full test. This method very much lessens any advantage that might go to one set of pupils from having seen the other set take the tests.

[7] The effect of second trials is explained in Chapter V, p. 48.

[8] All instructions for using the Scale of Motor Ability Tests are contained in Chapter VIII, p. 100.

CHAPTER III

THE VALIDITY OF THE TESTS AS MEASURES OF MOTOR ABILITY

THERE is no uniformity of opinion among physical educators as to just what constitutes motor ability. Various tests have been proposed.[1] While these tests may measure motor ability, no evidence is advanced to prove whether or not, or in what amounts, such is the case.

This condition is probably due, in part, to two apparently conflicting reasons. On the one hand, the scientific defining of native motor ability or abilities is rather difficult, while on the other hand, the subjective defining of motor ability appears rather easy.

Any observant youth or adult, and certainly a trained physical educator, can watch a number of participants in action in a variety of activities and select rather easily those individuals whom he feels have the most motor ability. He can also select those who have, as he feels, the least motor ability. The selection of those who are of about average ability is, of course, more difficult. This selection is, obviously, easier for the extremes than for the average group.

In forming this judgment the following considerations are doubtless among those mental criteria set up:

(1) Ease and proficiency in learning new coördinations.
(2) Proficiency in a wide variety of motor activities.
(3) Performing activities with easy and graceful form.

---

[1] "Motor Ability Tests," *American Physical Education Review,* December, 1924. Report of Committee of American Physical Education Association. Proceedings of Athletic Research Society, 1920-1921.
J. F. Williams, "Physical Education," pp. 272 ff.

A group of 75 physical educators ranked[2] these judgment criteria of motor ability as shown in Table II, page 14.

TABLE II

JUDGMENT CRITERIA OF GENERAL MOTOR ABILITY RANKED BY PHYSICAL EDUCATORS AS TO VALUE

CRITERIA OF GENERAL MOTOR ABILITY	Average Rank as to Value
1. Learning new activities easily	1.76
2. Skill in a variety of activities	2.02
3. Easy and graceful form in performance	2.41
4. Great ability in some special line	3.81

Table III shows how the group of physical educators used in Table II, page 14, ranked ten activities as to the reliability with which judgments of motor ability could be based upon skill in performance.

It will be apparent that the decision to have the motor ability tests consist of activities of the stunt nature was supported by the opinion of others. Tumbling and stunts involve an active manipulation of the whole body. All of the muscles are called into play. There must be good balance control and proper timing of effort. Considerable strength is also required.

All of the first four activities (Table III, page 15) call for a wide variety of muscular reactions. The place of dancing in Table III may be explained by the fact that the women in the group do considerable natural dancing, folk dancing and clog dancing. Natural dancing calls for much running and leaping, and general free body movements.

For purposes of this study, the term "motor ability" is

[2] Students majoring in physical education at Teachers College, Columbia University, 1925-1926—seniors and graduate students, men and women, nearly all of whom have taught.

TABLE III

RELATIVE VALUE OF TEN MOTOR ACTIVITIES, SKILL IN WHICH CAN BE TAKEN AS A CRITERION OF GENERAL MOTOR ABILITY

Motor Activity	Number of Judges Giving Ranks	Average Ranking
1. Tumbling and stunts....	74	2.5
2. Dancing	75	3.9
3. Track and field sports...	75	4.0
4. Apparatus exercises.....	71	4.3
5. Basketball	73	5.0
6. Swimming	75	5.8
7. Baseball	74	6.2
8. Field Hockey.....	75	7.4
9. Soccer football.....	71	7.5
10. Gymnastic Drills.....	73	8.7

used to apply to that ability which is more or less general, which is more or less inherent, and which permits an individual to learn motor skills easily and to become readily proficient in them. That facility in learning some athletic skills is not largely dependent upon intelligence is illustrated by the fact that the correlation coefficient [3] between scores on six basketball achievement tests [4] and scores on the Thorn-

[3] Using the Pearson product-moment formula, $r = \frac{\Sigma XY}{N\sigma x\sigma y}$ as discussed by H. O. Rugg, "Statistical Methods Applied to Education," pp. 251 ff. Unless stated otherwise, all coefficients of correlation used in this study are by the product-moment method.

[4] See Chapter VI, p. 76.

dike College Entrance Examination for a group of ninety college women [5] was $r = -.15$, P.E. $= .069$ [6]

Motor ability is not used here to designate achievement or proficiency in specific athletic skills, as is sometimes erroneously done.[7]

The Selection of Criteria for Determining the Validity of the Tests as Measures of Motor Ability

A perfectly valid test is one which measures exactly that which it is intended to measure.[8] If one wishes to construct a test of intelligence he has available as criteria, with which to validate his test, several standardized tests of intelligence. If one wishes to construct a test of ability to run, he has available previously established tests of running ability with which to compare his new test.

For motor ability, however, there are no scientifically prepared tests which can be used as a basis for comparison, *i.e.*, for validating new tests. Other criteria of validity must be found. The following possible criteria of validity suggest themselves:

(1) Judgment ratings on motor ability by qualified judges.
(2) Scores on athletic events.
(3) Proficiency in games of skill.

---

[5] Women seniors at Teachers College, majoring in physical education. The tests were given following a six-weeks' learning period.

[6] The Probable Error, P.E. $= .67449 \times \frac{1 - r^2}{\sqrt{n}}$, where r is the product-moment coefficient of correlation and n is the number of cases. It is a measure of unreliability and its relative size shows the limits outside of which it is very improbable that the true value will fall. Common statistical practice has set the criterion that the coefficient of correlation should be at least three times as large as the P.E. (H. O. Rugg, "Statistical Methods," p. 272 ff.)

The P.E. used in this study are computed from tables in K. J. Holzinger's "Statistical Tables for Students in Education and Psychology," pp. 60 ff. Those used in this study may be slightly in error in the third decimal place.

[7] Report of a National Committee on Motor Ability in the *American Physical Education Review*, "Motor Ability Tests," December, 1924, Vol. XXIX, No. 10.

For a discussion on the classification of tests in physical education, see an article under this name by the author, accepted for publication in the *American Physical Education Review*, January, 1927.

[8] William A. McCall, "How to Measure in Education," pp. 195 ff.

JUDGMENT RATINGS AS A CRITERION OF VALIDITY

The Motor Ability Test, consisting of 33 test elements, was given to 451 Junior and Senior High School boys and girls at Baldwin, Long Island,[9] during April, 1925. During the second week of April, 1925, a judgment score was given to each of the 451 pupils.[10]

For purposes of forming judgments as to the pupils' motor ability, the criteria of general motor ability explained on page 13 were used. These criteria were discussed with all the judges in an effort to have all get the same idea of what was meant by the term "motor ability."

Each child was given a judgment rating or score which represented the opinion of the judge as to its motor ability. The activities upon which this judgment was based were those found in the regular physical education program and on the play field, and not upon those found in the tests.

Each child was first placed in one of five groups lettered A, B, C, D, E, of which Group A was used for those believed to have superior motor ability. In addition to this rough grouping, a finer grading was provided by giving a number grade within each group. These number grades ran from 1 to 20. The number grades were so distributed as to assist in having the judgment ratings for the total group of children distributed somewhat in accordance with a normal distribution.[11]

Table IV shows what judgment ratings, or grades, were given and how they were distributed.

From Table IV, page 18, it will be seen that the judgment ratings were fairly well distributed. However, when the distribution of number ratings, or grades in each group, is ex-

[9] The tests were administered for the author by Carl Stockdale, Director of Physical Education, Baldwin, L. I.

[10] The judgment ratings were given by Mr. Stockdale, assisted by Miss E. Bedell, school nurse, and ten students, boys and girls, of superior athletic ability.

[11] This plan is based upon the assumption that motor ability is a trait present among individuals in accordance with a normal distribution.

TABLE IV

SHOWING HOW JUDGMENT RATINGS OF MOTOR ABILITY WERE TO BE GIVEN *

Degree of Motor Ability	Group Rating	Grades in Each Group	Per Cent Planned to Receive Group Ratings	Per Cent Actually Receiving Each Group Rating
Excellent	A	19, 20	10	3.5
Above Average..	B	16, 17, 18	15	11.3
Average	C	6, 7, 8, 9, 10, 11, 12, 13, 14, 15	50	59.2
Below Average.	D	3, 4, 5	15	21.1
Poor	E	1, 2	10	4.9

* For 451 Baldwin, L. I., pupils, 1925.

amined, it is seen that a multimodal distribution results, from which it would appear that twenty possible ranks proved more than could be handled efficiently. Chart I, page 131, shows the distribution of number judgment ratings.

The Motor Ability Tests were scored by simply counting the total number of tests passed.[12] The correlation coefficient between the test scores and the judgment ratings for the 451 Baldwin pupils was

$$r = .578,\ \text{P.E.} = .021,\ \text{P.I.}^{13} = .184,\ \eta yx^{14} = .638$$

Chart I, page 131, shows the scattergram from which this correlation was obtained. From inspection it might appear

[12] Other methods of scoring are explained in Chapter VI.

[13] Reference to the Predictive Index will be found on page 20, footnote †.

[14] The formula for the correlation ratio of x upon $y = \eta yx = \sqrt{1 - \frac{\sigma^2 ax}{\sigma^2 1}}$

T. L. Kelley, "Statistical Methods," pp. 238 ff.

doubtful as to whether or not the scattergram shows rectilinearity and, hence, that the product-moment formula for r should be used. Blakeman's test for linearity[15] produced an index of 1.133, which falls within the limits set for linearity.

At this point it was decided to drop three tests which appeared to call for an undue amount of judgment on the part of the scorer. The remaining 30 tests were arranged in two forms, A and B, which were equated on the basis of the difficulty of the tests as shown by the Baldwin data.[16] Further testing was carried on upon the basis of the 30 tests included in Forms A and B.

The correlation coefficient for the 30 tests in Forms A and B and the judgment ratings (ungrouped, *i.e.*, where the twenty ratings were used) was

$$r = .56$$
$$P.E. = .022$$
$$P.I. = .172$$

The correlation coefficients for the 231 boys and the 220 girls of the Baldwin group, for both grouped and ungrouped judgment ratings, are given in Table V, page 20. This table gives, also, the coefficient of correlation for the final scale tests with the judgment ratings, for all in the Baldwin group.

Since scores on several tests for the same individuals (together with criterion scores)[17] are available it should be possible to weight the tests or to make some combination of them that will yield higher correlations with the criterion than would otherwise be obtained.

Such procedure has been followed and will be reported in a later chapter.[18] It may suffice to indicate here that test scores were weighted on the bases of difficulty, the sum of criterion scores for tests passed, the size of the difference

---

[15] H. O. Rugg, op. cit., p. 283.

[16] An explanation of the procedures used in forming equivalent tests will be found in Chapter V, p. 63.

[17] Criterion scores were judgment ratings in this case.

[18] Chapter V, p. 48, Scaling the Tests.

TABLE V

CORRELATION COEFFICIENTS FOR SCORES ON THE MOTOR ABILITY TEST AND JUDGMENT RATINGS FOR BOYS AND GIRLS IN THE BALDWIN GROUP

Method of Handling Judgment Ratings	Pupils Ages 10-18	Number	Correlation * Coefficients for Scores and Judgments	P.E.	P.I.†
Grouped Judgments	Boys	231	.68	.025	.267
(5 groups)	Girls	220	.52	.034	.146
Ungrouped Judgments	Boys	231	.61	.029	.207
(Scale of 20)	Girls	220	.47	.038	.117
Ungrouped (Scale of 20)	Boys and Girls	451	58‡	.021	.186

* The 30 tests in Forms A and B, when the score equals the number of tests passed.

† The Predictive Index formula, P.I. $= 1 - \sqrt{1 - r^2}$ is discussed by E. M. Bailor, "Content and Form in Tests of Intelligence," pp. 27-28.

Predictive indices are comparable measures and should be used for comparisons rather than correlation coefficients.

‡ The tests used in obtaining this coefficient are the 20 tests of Forms M and N comprising the final Scale of Motor Ability Tests.

between criterion scores of tests passed and tests failed, and the size of the index of validity of each test.

The highest correlation obtained for tests with the criterion when weightings were used was *.60*, while the correlation of the selected scale tests with the criterion was *.58*.

Conclusion

It is, perhaps, reasonable to expect that the judges could form more accurate opinions about the motor ability of boys

than that of girls. Three of the possible reasons for this assumption are:

(1) The judges were better acquainted with the boys than with the girls.

(2) They were more accustomed to forming opinions about boys in boys' activities than about girls in girls' activities.

(3) They were less well acquainted with the motor activities of the girls' program of physical education.

The highest correlation coefficients are obtained when the grouped judgment ratings are used. Although five divisions are fewer than are advised for product-moment correlation coefficients,[19] the procedure may be justified here since the method of classifying pupils into five divisions was a first step in the procedure and is supported by findings based upon ratings on a scale of 20 divisions.

A coefficient of correlation of .68 is sufficiently high to indicate that there is marked evidence that motor ability as shown by the motor ability tests will tend to correspond with motor ability as estimated by judges. A product-moment correlation coefficient of .60 or above, for this type of data, may be regarded as a high coefficient.[20]

THE VALIDITY OF THE MOTOR ABILITY TESTS SHOWN BY CORRELATIONS WITH ACHIEVEMENT IN ATHLETIC EVENTS

Many teachers of physical education have believed that if individuals were allowed sufficient practice to obtain their best scores on a variety of athletic events, the aggregate total of these scores would be a fair index of motor ability. It is the purpose of this section to show that the motor ability tests do correlate highly with such data.

A part of the program of physical education for boys in Lincoln School, New York City, consists of obtaining the

[19] H. O. Rugg, op. cit., p. 84.
[20] *Ibid.*, p. 256.

best scores that each boy can make on a "Decathlon Test."[21] The events of this Decathlon Test are as follows:[22]

(1) Pull-ups (chinning).
(2) Fence Vault.
(3) Running High Jump.
(4) Potato Race.
(5) Rope Climb.
(6) Jump and Touch.
(7) Running Broad Jump.
(8) Dash (75 and 100 yards).
(9) Baseball Target Throw.
(10) Basketball goal throwing.

Scores in each event are transmuted into point scores in accordance with scales which are in use in the school, and a total point score is computed for each boy. This total score is obtained by simply adding the point scores earned in each event.

The Motor Ability Test was given to ninety boys of the Junior and Senior High School.[23] There was available, then, for each of the 90 boys, a score[24] on the Motor Ability Test and a Decathlon Test score. The correlation coefficient obtained was

$$r = .804$$
$$P.E. = .011$$
$$P.I. = .405$$

When it is recognized that this coefficient was obtained without resorting to statistical methods to weight the Motor Ability Test scores or the Decathlon Test scores in an optimum

[21] The motor ability tests and the decathlon tests, whose data is here referred to, were given to his pupils by Colba Gucker, Director of Physical Education, Lincoln School, New York City, in March, 1925 and March, 1926.

[22] These events, with one exception, have been used for at least three years in this school.

[23] April, 1925.

[24] The score consisted of the total number of tests passed on the first trial.

manner to produce the highest possible coefficient, its significance is perhaps more apparent.[25]

One year later the Motor Ability Tests were given again to 92 Lincoln School boys. At this time, only 6 Decathlon Tests had been given, namely, pull-up, fence vault, potato race, rope climb, high jump, and speed pass.[26] The coefficient of correlation for Motor Ability Test scores with these Decathlon scores was

$$r = .70$$
$$P.E. = .036$$
$$P.I. = .286$$

Using only the Motor Ability Tests included in the final scale, the coefficient of correlation was

$$r = .73$$
$$P.E. = .033$$
$$P.I. = .317$$

Conclusion

A possible criterion of motor ability might be the sum of scores on a variety of athletic events in which performers have been given ample opportunity to make the best scores of which they were capable. Decathlon Tests constitute a variety of such athletic events.

Two sets of data giving two coefficients of correlation for the Motor Ability Tests and a criterion consisting of Decathlon Tests have been presented. When the Decathlon Tests comprised 6 events of coefficient of correlation of .73 was obtained. When 10 events were used as a criterion, the coefficient of correlation was *.80*. These correlation coefficients were obtained without weighting the various elements so as to produce optimum coefficients.

It is possible that a criterion consisting of many more

[25] Methods of weighting test elements so as to produce an optimum scoring key are discussed in Chapter V, p. 48 ff.

[26] Not contained in the Decathlon Tests used in 1925.

varied sorts of athletic events might produce still higher coefficients of correlation with the Motor Ability Tests.

MOTOR ABILITY SCORES OF ATHLETIC TEAM MEMBERS

When there is free competition for membership on athletic teams, most of those students becoming team members may be expected to possess superior motor ability. If the tests proposed by this study really measure motor ability, it might be reasonable to expect that, on the average, scores of team members on the Motor Ability Tests would be higher than scores on the same tests for the student body as a whole. This has been found to be the case.

In the Baldwin group, 56 pupils, 34 boys and 22 girls, were indicated as being members of the school teams, some pupils being members of several teams during the year. The mean scores on the Motor Ability Tests [27] of team members compared with the student body as a whole was:

Mean score for team members = 25.893
Mean score for all students = 20.812

That the judgment ratings or criterion scores of team members would be, on the average, higher than those of the student body as a whole would also be expected. This was found to be the case:

Mean judgment rating for team members = 14.965
Mean judgment rating for all students = 10.361

The athletic teams were as follows: Football, boys' basketball, girls' basketball, baseball, and track and field (boys and girls).

Conclusion

A comparison of the mean scores for team members with the mean scores for all students on the Motor Ability Tests shows that, on the average, team members did considerably

[27] For the 30 tests of Forms A and B.

better in the tests than did the students as a whole. This superiority amounts to 23.5 per cent, thus tending to support the validity of the tests as measures of motor ability.

The validity of judgment ratings as a criterion of motor ability is slightly supported by the fact that the mean judgment rating for team members is 30.8 per cent higher than the mean judgment rating of all students combined.

SUMMARY

(1) The determination of the validity of tests, that is, the extent to which they measure what they are supposed to measure, is of fundamental importance. Tests whose validity is unsupported by objective evidence cannot be regarded seriously.

(2) That the types of activities chosen as tests have characteristics which appear to make them suitable means for distinguishing degrees of motor ability is supported by the judgments of others.

(3) As there have been no previous tests of motor ability whose validity has been established, the selection of a criterion with which to validate the proposed tests is a difficult matter. Three sorts of criteria have been studied. These are (a) judgment ratings, (b) scores on a variety of athletic events, and (c) achievement in athletic games.

(4) The use of judgment ratings as a criterion by which to establish validity is a questionable procedure. Such a criterion would not be satisfactory were proved criteria of an objective nature available. The absence of such criteria, however, warrants its use.

Coefficients of correlation for the Motor Ability Tests with judgment ratings range from *.47* to *.52* for girls to *.61* and *.68* for boys. The coefficient of correlation for the final scale tests with judgment ratings as the criterion was $r = .58$.

(5) A second type of criterion consists of scores on a variety of athletic events in which pupils have had opportunity to make their best records. Coefficients of correlation ob-

tained for scores on the Motor Ability Tests with scores on a variety of athletic events range from *.70* to *.80*. The largest coefficient was obtained when the largest number of events were included in the criterion.

This may indicate that the best criterion might consist of scores on 10 to 15 athletic events, scores in a variety of athletic game skills, and judgment ratings, combined in an optimum manner.

(6) Students who were members of school teams were shown, on the average, to have Motor Ability Test scores considerably above the average for students in general. The judgment ratings assigned team members were also considerably above the average.

(7) Evidence as to the validity of the tests as measures of motor ability is, perhaps, not as conclusive as that used in validating tests of intelligence, for example. It is, however, highly satisfactory in view of the complicated nature of the ability in question, and of the absence of better criteria.

CHAPTER IV

THE DETERMINATION OF RELIABILITY AND OBJECTIVITY

A TEST has perfect reliability when an identical examiner secures identical scores from two or more applications of the same test to the same pupils.[1] Educational tests rarely, if ever, have perfect reliability, but the higher the reliability the better.

Unreliability, or lack of agreement between repeated applications of the Motor Ability Tests by the same examiner to the same group of children, might arise from causes, among which may be mentioned:

(1) The equipment required by the test may vary between testings.
(2) The description of the tests and instructions for its administration may not be sufficiently clear and definite.
(3) The examiner may not be able to demonstrate the test.
(4) There may be large errors due to chance scores.
(5) There may be variation in the ability of the pupils.
(6) Incompetence or dishonesty on the part of children who score each other when the test is given as a group test or when the subject scores himself may cause unreliability.

Changes in Equipment Not a Cause of Unreliability

Since the only equipment required for the administration of the Motor Ability Test is a watch, pencils and printed

[1] William A. McCall, "How to Measure in Education, p. 307.

or mimeographed scoring blanks, there is practically no possibility for unreliability resulting from changes in equipment.

Instructions for Administering the Test Are Clear and Definite

The description of the tests was first worded as clearly as possible. The tests were then given by the writer[2] and changes made toward greater definiteness. In final form the instructions for administering the tests are definite and simple. The procedure is familiar to teachers of physical education, as are most of the activities comprising the tests. The description of the tests is accompanied by photographs of two positions for each test. The tests have been given by a number of different examiners with no difficulty. It would appear, therefore, that any unreliability entering from this score would be small.

The best explanation of the test is that which is accompanied by a demonstration. Very few examiners will be able to perform each of the test activities successfully. However, it is not essential that the test be demonstrated successfully. The purpose of the demonstration is to make clearer what constitutes success and what failure. Demonstrating or indicating faulty performance may, therefore, be as valuable, if not more valuable, than demonstrating successful performance. The test instructions indicate what demonstration should be made and how it should be used in pointing out how to score.

The tests were so selected that their proper performance would be readily apparent. Experience has shown that pupils have little difficulty in grasping what is wanted or in detecting whether or not the performance has been successful.[3]

This study advances the belief that the Motor Ability Test can be given as a group test in which pupils score each other, with results sufficiently reliable to warrant comparative use of class or group averages, and to give a fairly reliable indi-

[2] To 5th and 6th grade boys in Horace Mann School, New York City.
[3] See subsequent sections of this chapter.

cation of individual motor ability. Data in support of this position follows.

COMPARISON OF FAILURES ON A FIRST TRIAL WITH FAILURES ON A RETRIAL[4]

All pupils in a group of 451,[5] who failed a test on the first trial, were given a second trial on that test. Table VI, page 30, gives the frequencies with which the separate tests were failed on first trial and on retrial. It gives, also, a per cent of constancy, *i.e.,* the per cent of first trial failures which failed also on the second trial.

From the data of Table VI, page 30, it will be seen that certain tests failed on the first trial were practically always failed on the second trial. Other tests were nearly always passed on the second trial. In one sense, those tests showing least change between first and second trials may be thought of as being more reliable. This reliability may be due either to a tendency of the test to eliminate chance success, or to a quality of the test which prevents improvement, or to other causes.

As the completed Motor Ability Test is to be scored on first trial only, the per cent of constancy shown by a single test element is not very significant as a measure of the reliability of first trial scores.

The data of Table VI, page 30, is referred to in another section relating to the final selection of tests to comprise the scale.[6]

The Mean Net Difference Between First and Second Trials

Another method of showing the amount of change between scores made on different trials is to determine the mean net difference between the two series of scores.[7]

Table VII, page 32, gives data for the group of Teachers

[4] The retrial was given immediately following the first trial.
[5] Baldwin, L. I., 231 boys and 220 girls, ages 10 to 18, 1925.
[6] Chapter V, p. 59.
[7] William A. McCall, "How to Experiment in Education," pp. 112-114.

TABLE VI

FAILURES ON SUCCESSIVE TRIALS AND PER CENT OF CONSTANCY *

Test † Number	Frequency of Successes on First Trial	Frequency of Failures on First Trial	Frequency of Failures on Both 1st and 2nd Trials	Per Cent ‡ of Constancy
15	29	422	406	96.1
14	124	327	289	88.4
30	125	326	278	85.2
24	337	114	88	77.2
13	197	254	196	77.2
12	221	230	174	75.6
27	225	226	162	71.6
26	231	220	157	71.3
25	238	213	144	67.6
28	175	276	184	66.6
11	290	161	104	64.6
21	371	80	50	62.5
10	305	146	90	61.6
17	415	36	22	61.1
7	367	84	45	53.6
2	421	30	16	53.3
5	383	68	36	53.0
19	398	53	28	52.8
29	138	313	158	50.4
4	386	65	32	49.2
23	367	84	37	44.0
22	367	84	37	44.0
8	358	93	39	41.9
20	386	65	27	41.5
6	369	82	34	41.4
9	311	140	57	40.7
18	409	42	14	33.4
3	412	39	10	25.6
16	431	20	4	20.0
1	442	9	0	0

* Baldwin group of 451 pupils.

† These numbers correspond to the Form A-Form B tests described on p. 2.

‡ The per cent of failures on both first and second trials, of failures on first trial.

College women, professional students in physical education, relative to the mean net difference between scores made on two successive trials [8] of the Motor Ability Test.[9]

The mean net difference shown in Table VII page 32 indicates the amount of error or change, on the average, to be expected between successive trials of the 30 tests in Forms A and B given to the type of students used in this Table.

This net difference, or unreliability, of .884, corrected .625, is small and amounts to a possible error on the average of approximately 2 per cent.

Were the Motor Ability Test perfectly reliable, there would be no increase or decrease in the scores made on successive trials. Where a difference does occur, it might be due to several causes:[10]

(1) There may be a difference in difficulty between the two sets of tests.
(2) Practice on the first test may influence subsequent scores.
(3) Changes in instruction, or growth, may influence the scores.

Since the data of Table VII [11] was obtained from a retrial upon the identical tests used for the first trial, it is probable that the second trial tests were easier than the first trial. The amount of this difference, on the average, was *1.21* (tests passed), an improvement of 4 per cent. Due to the nature of the situation, this improvement is remarkably small. This reasoning may also be applied to the possible influence of practice, since practice is one cause operating to make the tests slightly easier on the second trial.

As regards the data of Table VII, the effect of changes in instruction, or growth, would be practically negligible, as the second trial immediately followed the first trial and was

[8] The second trial was given immediately following the first trial.
[9] The thirty tests of Forms A and B are used in this data. For a description of Forms A and B, see p. 2.
[10] William A. McCall, "How to Experiment in Education," p. 113.
[11] p. 32.

TABLE VII

DETERMINING APPROXIMATE RELIABILITY BY THE MEAN NET DIFFERENCE METHOD* APPLIED TO SUCCESSIVE TRIALS

Positive Differences (2d Trial Score Exceeds 1st Trial Score By)	Frequency	Frequency times Difference	Negative Differences (1st Trial Score Exceeds 2d Trial Score By)	Frequency	Freq. × Diff.
0	14		1	11	11
1	19	19	2	7	14
2	20	40	3	5	15
3	10	30	4		
4	1	4			
5	2	10			
6	1	6			
	Sum	109		Sum	40

$N = 90$

Mean difference—non algebraic $\frac{(109 + 40)}{90} = 1.65$

Mean difference—algebraic $\frac{(109 - 40)}{90} = .766$

Mean net difference (unreliability) $= 1.65 - .766 = .884$

True unreliability † $(.707 \times .884) = .625$

* William A. McCall, "How to Experiment in Education," pp. 112-116.
† *Ibid.*, p. 116.

given under identical circumstances as to administration, examiner, etc.[12]

The function of the mean net difference is to reduce the non-algebraic mean difference by the amount of the algebraic mean difference. Assuming that influences due to 1, 2, 3 of page 31 have operated equally upon all pupils, the mean net

[12] The tests were given by the writer.

difference serves to eliminate their effect. "The net difference is approximately pure unreliability." [13]

THE RELIABILITY OF PUPILS' SCORES

In order to show how reliably pupils could score each other when the administration of the Motor Ability Test was least suitable for objective and reliable scoring, the following experiment was tried:

A group of twenty-seven sixth grade boys [14] were each given a scoring blank and were formed in a large circle about two boys who were to take the tests. Each boy in the scoring group was instructed to watch the boy performing the test and to score him a success if he thought the test was passed, and a failure if, in his judgment, the test had been failed. In order to eliminate fatigue, the two subjects alternated in taking the test. However, for each test, twenty-seven boys were all scoring the performance of the same individual subject.

Customary instructions were given before each test was attempted by the subject, but all instructions or hints for proper scoring were eliminated after the test had been performed. Any questions asked by the scorers were answered by the statement, "Do not ask questions. Score the boy as to whether you think he did or did not pass the test." No undue stress was laid upon the explanations preceding the test, as the purpose was to see what would happen were the test rather poorly administered.

Table VIII shows the per cent of agreement in the scoring of a single performer, by a group of pupils. Further use of this data is made in selecting tests for the final scale.

Conclusion

The average per cent of agreement among the scores of a group of sixth grade pupils when all were scoring the per-

[13] William A. McCall, "How to Experiment in Education," p. 113.
[14] From Horace Mann School, New York.

TABLE VIII

PER CENT OF AGREEMENT IN GROUP SCORING
BY SIXTH GRADE BOYS

Test †	Number Scoring	Per Cent of Identical Scores	Test	Number Scoring	Per Cent of Identical Scores
1	25	100.0	23	27	88.8
3	25	100.0	8	25	88.0
22	27	100.0	2	25	88.0
15	27	100.0*	17	27	85.2
11	27	96.3	29	27	85.2
13	27	96.3*	27	27	81.5
4	25	96.0	24	27	81.5*
25	27	92.5*	10	26	80.8
14	27	92.5*	6	25	80.0
30	27	92.5*	18	27	77.7
5	25	92.0	7	25	76.0*
12	27	88.8*	26	27	74.1*
16	27	88.8	9	26	68.0*
19	27	88.8	20	27	66.6
21	27	88.8	28	27	59.9*
Mean		94.22	Mean		78.73

* Tests were failed.
† See page 2 for a description of these tests.

formance of a single individual, was approximately 86 per cent. Care was taken to make no suggestions after a performance had been made that could possibly influence the scoring of the test. Had suggestions as to scoring been given following each performance, the writer feels there would have been practically 100 per cent agreement on the scoring of each test. However, no data is available in direct support of this position. It should be noted that the pupils in this group are approximately of the lowest age for which the Motor Ability Test is recommended as a group test.

CORRELATIONS: SELF-CORRELATION BETWEEN TWO APPROXIMATELY EQUIVALENT FORMS OF THE MOTOR ABILITY TEST

Following the application of the tests to the Baldwin group,[15] 30 tests were selected and arranged into two forms, known as Form A and Form B. These forms were equated on the basis of the type of test and upon the difficulty of the tests as determined by the per cent of successes for each test.[16]

The correlation coefficient for Form A with Form B for the Baldwin group is

$$r = .71$$
$$P.E. = .0151\ ^{17}$$

If it is desired to raise this self-correlation coefficient to .9, the Spearman formula may be used. This formula as given by McCall,[18] is

$$N = \frac{rx - rirx}{ri - rirx}$$

where N is the number of tests requires to yield rx; rx is the desired (.9) self-correlation coefficient; and ri is the self-correlation coefficient of one form with another form of the test. Thus, the number of forms required to yield the desired self-correlation coefficient of .9 may be found by substituting in the formula the desired self-correlation coefficient of .9 for rx and the obtained self-correlation coefficient of .71 for ri, and solving for N

$$N = \frac{.9 - .71 \times (.9)}{.71 - .71 \times (.9)} = 3.67 \text{ or } 4$$

[15] Four hundred and fifty-one pupils, Baldwin, L. I., public schools.

[16] For further details relating to the formation of equivalent forms, see Chapter V, p. 48.

[17] See Table XI, p. 44 for self-correlation coefficients between other equated forms.

[18] William A. McCall, "How to Experiment in Education," p. 111.

This shows that if the mean of 4 equivalent tests is correlated with the mean of 4 other equivalent forms, the self-correlation coefficient resulting would be .9. This would require 8 equivalent forms in all.

THE MEAN NET DIFFERENCE BETWEEN FORM A AND FORM B

McCall has suggested that the self-correlation coefficient is, perhaps, not the most intelligible way of reporting reliability. The determination of the mean net difference between two forms of a test is suggested and has been presented elsewhere in this chapter.

Table IX, page 37, shows the computation of the mean net difference between Forms A and B.

Conclusion

The procedure illustrated in Table IX is satisfactory since the variability of scores on the two forms is practically the same.

The mean net difference, corrected, is 1.061 tests passed. Assuming that this difference might possibly be 30, the net difference of 1.061 may be thought of as being three per cent of unreliability.

THE MEAN NET DIFFERENCE BETWEEN TWO SETS OF MOTOR ABILITY TEST SCORES OVER AN INTERVENING YEAR

Table X, page 38, presents the mean net difference of the two series of scores covering the intervening year.

TABLE IX

DETERMINING THE UNRELIABILITY OR MEAN NET DIFFERENCE BETWEEN FORM A AND FORM B *[1]

Positive Differences Scores † on Form B Exceed Scores on Form A	Frequency	Frequency Times Difference	Negative Differences Scores on Form A Exceed Scores on Form B	Frequency	Freq. X Diff.
0	92				
1	85	85	1	89	89
2	52	104	2	53	106
3	20	60	3	27	81
4	13	52	4	10	40
5	4	20	5	2	10
6	3	18			
Sum		339	Sum		326

$$\text{Mean difference, non algebraic} = \frac{339 + 326}{451} = 1.474$$

$$\text{Mean difference, algebraic} = \frac{339 - 326}{451} = -.029$$

Mean net difference (unreliability) $= 1.474 + .029 = 1.503$

True unreliability ‡ $(.707 \times 1.503) = 1.061$

S.D. Form A $= 2.141$ S.D. Form B $= 2.247$ §

[1] For 451 pupils of the Baldwin group.

* For the tests included in these two forms, see Chapter I, p. 2.

† Scores equal the sum of tests passed.

‡ William A. McCall, "How to Experiment in Education," p. 116.

§ The formula for the standard deviation, S.D. $= \sqrt{\frac{\Sigma fd^2}{N}}$ is "the sq. root of the arithmetic mean of the sqs. of the deviations from the average of the distribution," H. O. Rugg, "Statistical Methods Applied to Education," p. 168. The computation of the S.D. by the short method is explained pp. 170-172.

TABLE X

THE MEAN NET DIFFERENCE BETWEEN TWO SETS OF MOTOR ABILITY TEST SCORES OVER AN INTERVENING YEAR *

Positive Differences 1926 Exceeds 1925	Frequency	Frequency Times Difference	Negative Differences 1925 Exceeds 1926	Frequency	Frequency Times Difference
0	14	0			
1	4	4	1	8	8
2	6	12	2	3	6
3	3	9	3	1	3
4	0	0	4	1	4
5	0	0	5	2	10
6	1	6			
7	1	7			
Sum		38	Sum		31

N = 44

Mean difference, non-algebraic $= \frac{38 + 31}{44} = 1.568$

Mean difference, algebraic $= \frac{38 - 31}{44} = .159$

Mean net difference (unreliability) = 1.568 — .159 = 1.409

True unreliability † = (.707 × 1.409) = .996

* Group of 44 college women.

† William A. McCall, "How to Experiment in Education, p. 116.

CORRELATION: THE COEFFICIENT OF CORRELATION FOR SCORES ON THE MOTOR ABILITY TESTS ONE YEAR WITH SCORES ON THE SAME TESTS AND SAME STUDENTS ONE YEAR LATER

Data for three groups of students will be presented (1) college women, (2) public school children and (3) private school children.

(1) College Women

A group of 44 women, professional students in physical education, took the Motor Ability Tests[19] in 1925 and repeated the same tests approximately one year later, in 1926.

The coefficient of correlation for their 1925 scores with their 1926 scores was

$$r = .66$$
$$P.E. = .058$$
$$P.I. = .249$$

(2) Public School Children[20]

106 pupils, ages 13, 14 and 15, of the Baldwin group, repeated the same tests[21] in 1926.[22] The coefficient of correlation for their 1925 scores with their 1926 scores was

$$r = .71$$
$$P.E. = .033$$
$$P.I. = .299$$

(3) Private School Children[23]

A group of 88 pupils of the Lincoln School group repeated, in 1926, the same tests[24] they had taken in 1925.[25] The coefficient of correlation for their 1925 scores with their 1926 scores was

$$r = .82$$
$$P.E. = .024$$
$$P.I. = .424$$

Conclusion

There is a highly marked tendency for individuals to make corresponding scores on the Motor Ability Tests between two

---

[19] Thirty tests of Forms A and B.
[20] Junior and Senior High School pupils.
[21] The final scale tests of Forms M and N are referred to here.
[22] Given by the same examiner.
[23] Junior and Senior High School pupils.
[24] The final scale tests of Forms M and N are referred to.
[25] Given by the same examiner both times.

administrations of the tests even though a period of one year intervenes between examinations, as shown by coefficients of correlation of $r = .66$, $r = .71$, and $r = .82$.

In each case, the intervening time included the summer months. Of the groups examined, the most homogeneous in point of age was that of the Baldwin group. The least homogeneous was that of the college women.

Despite variations of environment during the months intervening between testings, the data presented indicated that the Motor Ability Tests show a marked degree of reliability. Had the intervening period been shorter in point of time it is, perhaps, reasonable to expect that correlations even higher than those obtained would have resulted.

CORRELATION: FAILURES ON FIRST TRIAL WITH FAILURES ON SECOND TRIAL

The Baldwin group of 451 children were given a first trial on each test and immediately following this, all who failed were given a second trial. Those who passed on the first trial were not given a second trial. Table VI, page 30, gives the frequency of failures on each test for both trials.

One measure of the reliability of the tests would be the coefficient of correlation for first trial failures with repeated failures on the same test on a second trial. If the tests had absolute reliability, a test failed on the first trial would be failed also on the second trial. There would, therefore, be no difference between an individual's score on his first trial and his score on his second trial. To express the same thing in terms of failures, an individual would make the same number of failures on his second trial that he made on his first trial.

The degree of correspondence between the number of failures on a first trial and the number of failures on a second trial may be expressed by the coefficient of correlation for these scores. If there is perfect reliability, this coefficient would be 1.0. If there is perfect unreliability, or pure chance, this coefficient would be 0.

Chart II, page 132, shows the scattergram for this coefficient of correlation.[26]

The coefficient of correlation computed from the data of Chart II was

$$r = .90$$
$$P.E. = .006$$
$$P.I. = .564$$

Conclusion

It must be noticed that in the data presented, only the first trial failures were given second trials. There was, therefore, a certain restriction upon the variation that could result on a second trial. However, as the mean of failures on the first trial for the Baldwin group of 451 pupils was approximately 10, and the mean number of failures on the second trial was approximately 7, there was ample chance for a great lack of correspondence between the two sets of scores. The coefficient of correlation obtained for this data was $r = .90$. This coefficient indicates a high degree of reliability.

CORRELATION: FIRST TRIAL WITH SECOND TRIAL SCORES MADE BY COLLEGE WOMEN

In order to show the reliability of the Motor Ability Tests as applied to a very different group of individuals from the Baldwin group, data relating to a group of 90 college women will be presented.

The 30 tests of Forms A and B were given. Second trials were allowed on all tests regardless of whether or not the test had been passed on the first trial. Second trials followed immediately after first trials.

The coefficient of correlation for first trial scores with second trial scores was

$$r = .73$$
$$P.E. = .033$$
$$P.I. = .316$$

[26] H. O. Rugg., op. cit., pp. 263-265, gives the steps for the computation of the product-moment coefficient of correlation from such a scattergram.

Conclusion

The tests were all given by the writer, who noticed a slight tendency on the part of students who had passed a test on the first trial to relax the attention and effort with which they attempted the second trial. This was partly due to the fact that no special caution had been given them that the second trial would count as well as the first trial.

The effect of this lessened effort to score a success on tests which they had already passed appeared upon those tests involving balance. To see what effect this might have on the correlation between first and second trials, 6 tests involving balance were eliminated and the coefficient of correlation was again obtained for first trial scores with second trial scores. This coefficient was

$$r = .78$$
$$P.\ E. = .028$$
$$P.I. = .374$$

That this coefficient was higher than was obtained when all tests were considered may have been due to a lessened effort on the second trial of tests already passed on the first trial, or to some other cause.

In any case, either coefficient indicates a fair degree of reliability.

OBJECTIVITY

In scientific measurements, objectivity is a key word. One of the most important reasons for developing scales to measure abilities and educational achievement is to secure tools which will help to eliminate subjective judgments.

The teacher of physical education must measure the ability of his pupils to perform, and to improve in, the skills he is attempting to teach. Heretofore, he has had practically no tools which would assist him in making such objective measurements, and has had to fall back upon subjective means, *i.e.*, judgments. One of the purposes of this study is to furnish

a tool to assist the teacher in making more objective measurements.

Tests are perfectly objective when two applications of the same test to the same pupils yield identical results. No educational tests are perfectly objective,[27] for objectivity, like reliability, is a matter of degree.

Objectivity is closely related to reliability, and, like reliability, depends in a large measure upon the clarity and definiteness of the instructions for administering the test. When two examiners give the same test to the same pupils, the scores resulting may differ because of two sorts of variables. One of these may be called the personal equation, or differences between the examiners and their understanding and skill in giving the tests. The other variable is the changes that occur in the reactions of pupils to two administrations of the test, *i.e.*, the reliability of the test.

When the Motor Ability Test is given as an individual test, its objectivity relates only to the performance of the examiner. The tests are so simple and easy to understand, and the instructions are so definite that the ordinary teacher of physical education should be able to follow directions exactly.

When the Motor Ability Test is given as a group test and pupils score each other, another element of personal variation may enter. This is the personal equation of the pupils.

Table XI, page 44, gives coefficients of correlation which furnish a partial check upon this phase of objectivity. The coefficient presented as number 1 in this Table was obtained from scores which include some tests for which pupil assistants acted as examiners.

The most economical way of administering the Motor Ability Test in point of time is to have the pupils score themselves. Data relating to this method of using the test is insufficient to establish satisfactory reliability and objectivity.

However, data is available for one group of 5th and 6th

[27] William A. McCall, "How to Measure in Education," pp. 311-313, gives a discussion of objectivity.

TABLE XI

CORRELATIONS INDICATING OBJECTIVITY AS RELATES TO GROUP SCORING

Data	Self-R
1. First trial failures with second trial failures *	.90
2. Form A with Form B †	.71
3. Form A with Form B ‡	.65
4. First trial score with second trial score §	.78
5. College women, one year with the next year ¶	.66
6. Scores in 1925 with scores one year later for the same pupils ‖	.82
7. Scores in 1925 with scores one year later for the same pupils **	.71

* Baldwin, 451 pupils, elementary, junior and senior high schools.
† Students in the above group.
‡ Thirteen-year-old pupils, 100, in Baldwin group.
§ College women, Teachers College, 90 students.
¶ College women, Teachers College, 44 Students.
‖ Lincoln School, 88 students, one year intervening between the two testings.
** Baldwin group, 106, one year intervening between the two testings.

grade boys, who scored themselves and at the same time were scored by the writer. These pupils were taken in small groups of four to six. Pupils were placed at one side of the room with their score sheets in front of them on the floor. The examiner had corresponding score sheets in a row ten feet away. The tests were explained and demonstrated as usual. Pupils were then instructed to perform each test and score their own efforts "just as you think you did the test." The boys were not allowed to see the examiner's score sheets.

In this method of scoring the tests, there was the usual

possibility of the pupils failing to understand the test, and in addition the opportunity to score the tests dishonestly.

The coefficient of correlation for scores on the Motor Ability Test [28] scored by the pupil taking the test with scores recorded by an examiner was:

$$r = .72$$
$$P.E. = .042$$
$$P.I. = .308$$

The mean score for tests passed scored by pupils themselves was 9.617, while the mean score for tests passed scored by the examiner was 10.00.

Conclusion

The number of pupils used, and the number of tests given, is too small to make the findings wholly significant. The pupils were fifth and sixth grade boys, ages principally 9 to 12, and represent the youngest pupils for which this method of scoring is suggested. The obtained coefficient of .72 should be viewed in this light.

Although it might be expected that pupils whose scoring was not watched at all might tend to favor themselves, a comparison of the mean scores shows that, as a matter of fact, they tended to mark themselves more severely than did the examiner.

SUMMARY

(1) Owing to the nature of the Motor Ability Tests proposed and to the method of group scoring, there may be some question as to the reliability and objectivity of the tests.

(2) The simplicity of the score form and the absence of any equipment eliminates equipment as a source of unreliability.

(3) The definiteness of instructions,[29] and the simplicity

[28] Form A only.

[29] Chapter VIII, p. 100 ff.

of the tests caused by the fact that they are scored either "success" or "failure" favors satisfactory reliability and objectivity.

(4) The mean net difference between first and second trials for a group of college women was small, .625.

(5) Reliability of the group method of scoring was shown, in part, by the average per cent of agreement resulting when a single subject was scored simultaneously by a group of young boys. This was 86 per cent for 30 tests, even though customary suggestions for scoring after each performance were omitted.

(6) Two equivalent forms of the Motor Ability Test produced a self-correlation coefficient of $r = .71$ for a group of 451 pupils, ages 10 to 18 years.

(7) The mean net difference between Form A and Form B was 1.061.

(8) The mean net difference between tests separated by ten months was low, namely, .996. It might be that this would be even lower for tests given closer together in time.

(9) The self-correlation between tests separated in time by approximately one year, for a group of college women, was $r = .66$.

(10) The self-correlation between tests separated by approximately one year, for elementary and high school pupils, was found to be $r = .71$ for one group, and $r = .82$, for another group.

(11) The self-correlation coefficient between successive tests, one immediately following the other, where only failures were scored, was $r = .90$. The self-correlation coefficient between successive trials for college women was $r = .73$.

(12) The objectivity of the Motor Ability Tests was partly established by its relation to reliability as expressed by self-correlations of

$$r = .65,\ r = .71,\ r = .73,\ r = .78 \text{ and } r = .90$$

(13) Objectivity was further established by a coefficient of correlation of $r = .72$, for self-scored tests by young pupils,

with the identical tests scored by an examiner. The difference between the mean of self-scored tests and the mean of examiner-scored tests showed a slight tendency on the part of pupils to score themselves more severely than did the examiner.

CHAPTER V

SCALING THE TESTS

It is the purpose of this chapter to show what data and what methods were used in scaling the tests. The problems presented for solution are:

(1) The determination of the optimum scoring key.
(2) The selection of tests to be included in the final scale.
(3) Formation of equivalent forms of the Test.
(4) The selection and application of a scaling technique.

The selection of tests comprising the final scale is based, mainly, upon data from the Baldwin group of 451 pupils, from grade six through the high school. Each of these pupils was given 33 tests. An immediate retrial was given on each test failed. It was decided, at this point, to eliminate three tests[1] thus leaving data on 30 tests.

DETERMINATION OF THE OPTIMUM SCORING KEY

The simplest scoring key is obtained by totaling the tests passed and calling the score made on the tests the sum of tests passed. Applying this method to the Baldwin data, the scores on the thirty tests used were found to range from 5 to 29, with the mean score being $M = 20.814$, and the Standard Deviation being $S.D. = 4.039$.

The coefficient of correlation for scores on these thirty tests with the criterion[2] was

$$r = .567$$
$$P.E. = .021$$

[1] These tests involved too much judgment in scoring. They are not described in this volume.

[2] The criterion here was the judgment ratings referred to in Chapter III, p. 20.

Weighting Tests in Accordance with Their Difficulty [3, 4]

It might be argued that those tests which proved most difficult should have a larger share in the final score. Although difficulty is not the best basis upon which to weight test elements, the results of applying this method are reported.

Table XII gives the per cent of pupils passing each test, and the weight which was accordingly assigned each test.

The 451 test blanks of the Baldwin group were scored when the tests were weighted in accordance with the data of Table XII, page 50. The tests thus weighted and correlated with the judgment ratings used as a criterion produced a coefficient of correlation of

$$r = .55$$
$$P.E. = .022$$

Weighting Tests on the Basis of Mean Criterion Scores

The criterion scores here considered are the judgment ratings obtained for pupils in the Baldwin group. It is possible therefore to find the sum of the judgment ratings, or criterion scores, of all the pupils who passed each test. From this data the mean criterion scores of the pupils passing each test can be computed.[5]

If the mean criterion score for Test 30 is found to be higher than the mean criterion score for Test 1, it is known that, on the average, pupils passing Test 30 are judged to have better motor ability than pupils passing Test 1.

When the tests are weighted in accordance with the size of the mean criterion score for each test, the coefficient of correlation for test scores with criterion scores should be higher than that obtained by using unweighted scores. Table

[3] T. L. Kelly, Statistical Methods, pp. 67-68.

[4] Wm. A. McCall, "How to Measure in Education," pp. 30-2.

[5] For an illustration of the method used in computing the sum of criterion scores, see Table XV, p. 56.

TABLE XII

Tests Weighted In Accordance with Difficulty

Test *	N	Pupils Passing Each Test	Per Cent Passed (Difficulty)	Weights Based on Per Cent Passed
1	451	443	98.20	1
16	451	433	96.00	1
2	451	422	93.57	1
17	450	414	92.00	1
3	451	414	91.79	1
18	451	409	90.87	1
19	451	398	88.24	1
4	447	388	86.82	1
20	450	389	86.44	1
5	450	383	85.11	1
6	451	375	83.14	2
21	451	371	82.26	2
23	451	369	81.81	2
22	451	368	81.59	2
7	451	365	80.93	2
8	451	360	79.82	2
24	450	335	74.44	2
9	451	309	68.52	2
10	451	307	68.07	2
11	451	296	65.63	2
25	449	239	53.23	2
26	451	231	51.21	2
27	451	226	20.11	3
12	450	220	48.88	3
13	450	196	43.55	3
28	451	177	39.24	3
29	450	136	30.22	3
30	451	129	28.60	3
14	451	122	27.05	3
15	451	31	6.87	4

* Numbers correspond with those of Forms A and B. Form A—Tests 1 to 15, Form B—Tests 16-30.

XIII, page 51, gives the mean criterion scores for each of the thirty tests of Forms A and B.

Recognizing that, other considerations being equal, it is an advantage to have a test that can be given in the shortest possible time, 10 of the 30 tests in Forms A and B were eliminated. The basis upon which this elimination was made will be discussed later in this Chapter.[6] These 20 selected

TABLE XIII

MEAN CRITERION SCORES FOR TESTS PASSED*

Test	Frequency Passed	Sum of Criterion Scores	Mean Criterion Scores	Weight†	Test	Frequency Passed	Sum of Criterion Scores	Mean Criterion Scores	Weight†
16	433	4285	9.89	.....	7	365	3763	10.31	1
1	443	4386	9.90	.....	23	369	3824	10.36	1
18	409	4108	10.04	.....	24	335	3525	10.52	1
21	371	3728	10.04	.....	25	239	2526	10.51	1
3	414	4165	10.06	.....	11	296	3204	10.82	2
9	309	3125	10.11	.....	28	177	1923	10.86	2
22	368	3722	10.11	.....	26	231	2532	10.96	2
17	414	4187	10.11	1	10	307	3365	10.96	2
20	389	3939	10.12	.....	12	220	2454	11.15	3
2	422	4277	10.13	.....	13	196	2212	11.29	3
6	375	3814	10.17	.....	29	136	1575	11.58	4
4	388	3985	10.27	1	30	129	1494	11.58	4
8	360	3700	10.28	1	27	226	2655	11.75	5
5	314	3224	10.29	1	14	122	1452	11.92	6
19	398	4098	10.30	1	15	31	408	13.16	8

* The 30 tests of Forms A and B are described on page 2 ff.
† Weights were assigned only to the 20 tests of Forms M and N.

tests were given the weights shown in Table XIII, page 51. The 20 tests were arranged into two approximately equivalent forms, called M and N.[7]

The coefficient of correlation for the 20 tests of Forms M and N weighted in accordance with the size of the mean

[6] Page 59 ff.
[7] r for Form M with Form N is r = .61, P.E. = .02.

criterion score of tests passed with judgment ratings as the criterion, is

$$r = .55$$
$$P.E. = .022$$

When the 20 tests of Forms M and N used above were not weighted, the coefficient of correlation with the criterion was found to be slightly higher.

$$r = .58$$
$$P.E. = .021$$

Conclusion

In an attempt to see the effect of reducing the number of tests and weighting those remaining on the basis of the mean criterion scores of tests passed a lower coefficient of correlation was obtained when tests were weighted than was obtained when the same tests were given no weights.

This result may indicate that the weights assigned were not the best that could have been selected. It may also indicate that when selected in accordance with their validity, it may be unnecessary to assign weights.

Weighting the Tests on the Basis of Difference Between the Mean Criterion Scores for Successes and the Mean Criterion Scores for Failures

In the plan of weighting just reported, only the mean criterion scores for tests passed were used. This plan does not involve the criterion scores of tests failed.

The plan to be reported in this section was devised to weight scores on a basis which would consider the mean criterion scores of tests passed and failed.[8] All of the 30 tests of Forms A and B were used.

The data used in this plan is presented in Table XIV, page 53.

[8] Explanation of the method used in computing the sum of criterion scores may be obtained from Table XV, p. 56.

The 30 tests of Forms A and B, when weighted as in Table XIV, page 53, were called Forms *R* and *S*. When scores on Form R plus Form S were correlated with the judgment ratings, the coefficient of correlation obtained was

$$r = .59$$
$$P.E. = .020$$

TABLE XIV

MEAN CRITERION SCORES FOR TESTS PASSED AND FOR TESTS FAILED*

Test	Mean Criterion Score		Difference Between Means	Weight Given	Test	Mean Criterion Score		Difference Between Means	Weight Given
	Tests Passed	Tests Failed				Tests Passed	Tests Failed		
16	9.90	9.11	.79	1	30	11.58	9.18	2.40	7
9	10.11	9.32	.79	1	29	11.58	9.12	2.46	4
21	10.05	9.01	1.04	2	12	11.15	8.65	2.50	5
22	10.11	8.76	1.35	2	24	10.52	7.91	2.61	5
13	11.29	8.76	1.53	3	23	10.36	7.62	2.74	5
25	10.57	9.04	1.53	3	11	10.82	8.03	2.79	5
28	10.86	9.22	1.64	3	5	10.29	7.49	2.80	5
6	10.17	8.35	1.82	3	14	11.92	9.10	2.82	5
18	10.04	8.12	1.92	3	17	10.11	7.05	3.06	6
1	9.90	7.87	2.03	4	4	10.27	7.12	3.15	6
8	10.28	8.23	2.05	4	10	10.96	7.53	3.43	6
26	10.96	8.71	2.25	4	15	13.16	9.62	3.54	7
7	10.31	7.98	2.33	4	19	10.29	6.62	3.67	7
20	10.13	7.78	2.35	4	27	11.75	7.97	3.78	7
3	10.06	7.67	2.39	4	2	10.13	5.93	4.20	8

* For 30 tests of Forms A and B, listed in order of size of difference between mean criterion scores. As weighted here, Forms A and B were called Forms R and S.

Conclusion

It was desired to weight test scores by a method which would consider the size of the mean criterion score for pupils passing each test, and also the mean criterion score for pupils failing each test.

One method of doing this would have been to give each test two weightings, one based upon the mean criterion score

of pupils passing the test and the other based upon the mean criterion score of pupils failing the test.

The method adopted was thought to accomplish the same purpose as each test was weighted on the basis of the difference between the mean criterion score of pupils passing and the mean criterion score of pupils failing each test.

This method was an improvement over the method of weighting on the basis of difficulty. Its merit cannot be compared with that in which weighting was on the basis of the mean criterion score of tests passed, as the latter method was applied to only 20 tests.

Weighting Scores on the Basis of the Validity of Each Test

By "validity" of a test is meant the amount of correlation existing between the test and the criterion. If it were possible to find the amount of this validity it would be possible to weight tests in a manner to produce an optimum coefficient of correlation between scores made on the tests and the criterion. Such an optimum weighting would enhance both the validity and the reliability of the Scale of Motor Ability Tests.

Each test used in this research can be scored in only two ways. It is either passed successfully, or it is failed. This makes it difficult to get any highly satisfactory statistical measure of correlation between scores made on a single test and the criterion.

The problem is that of finding a method which will show the relation of a variable of two categories with one of many. Two of the methods previously used have been (1) biserial correlation,[9] and (2) finding the percentage of the failing group whose criterion scores overlapped or exceeded the median criterion score of the passing group.[10]

A new method known as the McCall-Long Method[11] has

---

[9] T. L. Kelly, "Statistical Methods, pp. 245-249.

[10] E. Leona Vincent, "Study of Intelligence Test Elements," Teachers College, Columbia University, 1924.

[11] W. A. McCall, and his students, "Construction of the Multi-Mental Scale," Teachers College Record, Bur. Vol. XXVII, No. 5, January, 1926.

been tried by the writer in this study and will be reported here.

The McCall-Long method is an empirical method and, as yet, has no statistical measures of reliability. It is recommended by the authors as a method which has produced higher coefficients of correlation than obtained on identical data with other methods.

This method involves the following data for each test item:

(1) The sum of the criterion scores for all pupils passing the test.

(2) The sum of the criterion scores for all pupils failing the test.

(3) The number of pupils passing the test.

(4) The number of pupils failing the test.

(5) The mean criterion score for the successes (X).

(6) The mean criterion for the failures (O).

(7) The difference between the two means obtained from steps 5 and 6 above.

(8) The product of the frequencies of the successes (X) times the failures (O).

(9) The product of frequencies of successes times frequencies of failures times the difference between the mean criterion scores for success and failure.

(10) The total number of pupils making a score on the test.

(11) The quotient obtained by dividing the product of frequencies of successes times frequencies of failures times the difference between the mean criterion scores by the total number making some score.

$$\text{Step } 11 = \text{Step } 9 \div \text{Step } 10$$

From the procedure outlined, a single figure will be obtained, which for this data ranged from 13.5650 to 425.5727. The larger this number, the greater the validity of the test item as determined by this technique.

Table XV, page 56, illustrates the process of determining the sum of the criterion scores.

TABLE XV

ILLUSTRATING THE METHOD FOR DETERMINING THE SUM OF CRITERION SCORES FOR A SINGLE TEST ITEM *

Pupils	Successes (X)	Failures (O)
1	10	
2	16	
3		4
4	9	
5		12
6		1
7		10
8	18	
450	20	
451		8
N = 451	Sum = 73 N = 5	Sum = 35 N = 5
	Mean = 14.6	Mean = 7.0

* This process was used for each of the 30 test items for the Baldwin group of 451 pupils.

The steps in computing the validity of each test by the McCall-Long Method enumerated on page 55 are embodied in Table XVI, page 57, which will illustrate the process for Test Number *27*, the one having the highest validity.

The validity index obtained for each test is given in Table XVII, page 58, in which the 30 tests of Forms A and B are arranged in order of the size of this index. The Table gives, also, the data from which the validity of each test was determined, and the weights which were given each test in accordance with the size of its validity index.

When the 30 tests of Forms A and B were weighted in accordance with the size of their validity index, as shown

TABLE XVI

ILLUSTRATING THE DETERMINATION OF THE VALIDITY OF A SINGLE TEST ITEM *

Step Number		Success (X)	Failure (O)
1–2.	Sum of criterion scores per mark	2655	1794
3–4.	Frequencies per mark	226	225
5–6.	Mean criterion score per mark	11.7478	7.9733
7.	Difference between the means (11.7478 — 7.9733)	= 3.7745	
8.	Product of frequencies per mark (226 × 225)	= 50,850	
9.	Product of the difference between mean criterion scores and the product of frequencies	= 191,933.325	
10.	Total frequencies (Step 3 + 4)	= 451	
11.	Product of frequencies per mark times the difference between the mean criterion scores for each mark divided by total frequencies (Step 9 ÷ Step 10) 191933 ÷ 451	= 425.5727	
12.	Validity of the test	= 425.5727	

* This process was repeated for each test.

in Table XVII, page 58, and the resulting scores correlated with the criterion, the coefficient of correlation obtained was

$$r = .60$$
$$P.E. = .020$$

Conclusion

An empirical method was used to determine an index of the validity of each test. The coefficient of correlation obtained by the use of this method was $r = .60$, the highest obtained by any of the methods of weighting used in this study.

It should be remembered that the criterion used here consisted of judgment ratings as to general motor ability. Cor-

TABLE XVII

Data Relating to the Determination of the Validity of Each Test

Test No.	N.	For X Scores			For O Scores			Validity Index	Wgts.
		Frequency	Sum of Criterion Scores	Mean Score Per X	Frequency	Sum of Criterion Scores	Mean Score Per O		
16	451	433	4285	9.8961	18	164	9.1111	13.565	1
1	451	443	4386	9.9006	8	63	7.8750	15.9173	1
21	451	371	3728	10.048	80	721	9.013	68.113	1
18	451	409	4108	10.0440	42	341	8.1190	73.3207	1
9	451	309	3125	10.1132	142	1324	9.3239	76.7913	2
3	451	414	4165	10.0603	37	284	7.6756	80.9951	2
22	451	368	3722	10.1141	83	727	8.7590	91.7674	2
17	450	414	4187	10.1135	36	254	7.0555	101.2809	2
15	451	31	408	13.1613	420	4041	9.6214	102.194	2
2	451	422	4277	10.1350	29	172	5.9310	104.7551	2
6	451	375	3814	10.1706	76	635	8.3552	108.4011	2
20	450	389	3939	10.1259	64	498	7.7812	129.7358	3
8	451	360	3700	10.2777	91	749	8.2307	148.6910	3
5	450	383	3943	10.2950	67	502	7.4925	159.7824	3
4	447	388	3985	10.2706	59	420	7.1186	161.4218	3
7	451	365	3763	10.3096	86	686	7.9767	162.3766	3
13	450	196	2212	11.2857	254	2225	8.7598	168.8231	3
25	449	239	2526	10.569	210	1898	9.038	171.1378	3
19	451	398	4098	10.2965	53	351	6.6226	171.8389	3
28	451	177	1923	10.8644	274	2526	9.2189	176.9478	4
23	451	369	3824	10.3631	82	625	7.6219	183.8961	4
30	451	129	1494	11.5814	322	2955	9.1770	221.450	4
24	450	335	3525	10.5223	115	910	7.9130	223.1628	4
29	450	136	1575	11.5808	314	2864	9.1210	233.4299	5
14	451	122	1454	11.9188	329	2995	9.1033	250.5732	5
26	451	231	2532	10.9610	220	1917	8.7136	253.2436	5
12	450	220	2454	11.1545	230	1989	8.6478	281.8644	6
11	451	296	3204	10.8243	155	1245	8.0322	284.0389	6
10	451	307	3365	10.9609	144	1084	7.5277	336.5101	7
27	451	226	2655	11.7478	225	1794	7.9733	425.5727	8

relations with other criteria are reported in Chapter III, page 13.

The writer makes no defense of the method of determining the validity of each test other than to report the correlation coefficient resulting from its use.

It will be noticed that the McCall-Long method tends to produce the highest validity index when the frequencies per mark are most nearly equal. This would appear to involve an assumption that a test passed and failed by approximately

equal per cents of a group has greater discriminating power than a test for which there is a greater difference between the number passing and the number failing.

The question as to whether or not the data presented warrants weighting test scores on the basis of the validity index for each test presents itself.

The method of simply totalling the number of tests passed to obtain a pupil's score on the Motor Ability Test is by far the most economical procedure in point of time. This method eliminates the use of a scoring key and thus reduces by one the number of steps that must be taken to obtain a pupil's score on the whole test.

This saving in time, however, would not be justified if the use of a scoring key produced a score having considerably more validity.

The relative difference between the $r = .56$ for unweighted scores, and the $r = .60$, for scores weighted on the basis of the size of the validity index may be obtained by comparing the P.I., Predictive Index[12] for the two coefficients. These coefficients and their Predictive Indices are as follows:

$r = .56$, P.I. $= .175$ (unweighted scores)
$r = .60$, P.I. $= .200$ (weighted scores)

The difference of .025 may be thought of as representing the per cent difference between the two coefficients of correlation. This difference of 2.5 per cent is small.

The findings of this study seem to indicate that a scoring key in which each test passed is given a value of 1, the total score then being the sum of tests passed, is satisfactory.[13]

SELECTION OF TESTS TO BE INCLUDED IN THE FINAL SCALE

The advisability of securing a Scale of Motor Ability Tests which requires the least time to administer is apparent. It

[12] The formula for the Predictive Index, P.I. $= 1 - \sqrt{1 - r^2}$ is discussed by E. M. Bailor, "Content and Form in Tests of Intelligence," pp. 27-28.

[13] T. L. Kelly, "Statistical Methods," p. 67-68, points out that real weighting occurs even though no assigned weights are used.

was partly this consideration which turned this study in the direction of tests to be scored only on success or failure and which permitted use as a group test.

Limits to which the effort to get a short test may go are set by the necessity of establishing validity and reliability, to mention only two of the most important considerations.

The study of Intelligence Test elements has shown that the goodness of a test element is related to its difficulty and to its complexity. It is shown that complexity and difficulty tend to accompany one another but that they are not the same thing. It is seen, too, that goodness [14] is not peculiar to the most complex elements, but that some simple elements occasionally have more power of discrimination.[15]

It is the feeling of the writer, largely unsupported, that motor ability test elements will prove to display many of the same characteristics shown by intelligence test elements. It will probably be impossible to get a single test element which will adequately measure motor ability.

Data has been given in this chapter which tends to show that any reduction in the number of tests is apt to cause a lowered correlation with the criterion. This is partially shown by the following Table.[16]

The final reduction of the number of tests to the 20 included in Forms M and N was based upon the following considerations:

(1) The validity of the test elements.[17]
(2) The reliability of pupils' scores on each test element.[18]
(3) The per cent of constancy between successive trials.[19]

---

[14] By goodness is meant the worth of the element for discriminating between the intelligent and the unintelligent.

[15] E. Leona Vincent, "A Study of Intelligence Test Elements," pp. 34-35.

[16] Table XVIII, p. 61.

[17] Data given in Table XVII, p. 58.

[18] Data given in Table VIII, p. 34.

[19] Data given in Table VI, p. 30.

TABLE XVIII

COEFFICIENTS OF CORRELATION
DEPENDING UPON THE NUMBER OF TEST ELEMENTS USED

Test Elements	Number	r with Criterion *
Original elements (tests)	33	.58
Forms C and D †	12	.46
Forms E and F ‡	18	.57
Forms A and B	30	.55
Forms M and N §	20	.58

* Consisted of judgment ratings for the Baldwin group.
† Tests of balance primarily.
‡ Tests other than those of balance.
§ Tests selected on the basis of validity indices and other considerations.

(4) The size of the mean criterion score for successes on each test.[20]

(5) The difficulty of each test.[21]

Table XIX, page 62, contains the collected data relating to the five items listed above which were considered in the determination of which tests to eliminate.

The writer arbitrarily assigned a weight to each of the five items in accordance with his opinion of its importance for the task in hand. The number at the top of each column, in Table XIX, represents the weight given.

As shown by the above Table, 10 tests were eliminated. The resulting 20 tests were arranged into two approximately equivalent forms known as M and N. These two forms were equated on the basis of the size of the validity index.[22]

[20] Data given in Table XIII, p. 51.
[21] Data given in Table XII, p. 50.
[22] Explained in Chapter V, p. 54 ff.

TABLE XIX

DATA USED IN ELIMINATING CERTAIN TESTS FOR THE FINAL SCALE

Test Forms A and B	5 Validity*	Wt.	4 Mean Criterion Score†	Wt.	3 Percent of Constancy‡	Wt.	2 Percent of Identical Scores§	Wt.	1 Difficulty Percent Passed¶	Wt.	Total Weights‖
16	13.57	5	9.90	4	20.0	3	88.8	0	96.00	1	13
1	15.92	5	9.90	4	0	3	100.0	0	98.20	1	13
21	68.11	5	10.05	4	62.5	1	88.8	0	82.26	1	11
18	73.32	5	10.04	4	33.4	3	77.7	2	90.87	1	15
9	76.79	5	10.11	4	40.7	3	68.0	2	68.52	0	14
3	80.99	5	10.06	4	25.6	3	100.0	0	91.79	1	13
22	91.77	5	10.11	4	44.0	3	100.0	0	81.59	1	13
6	108.40	5	10.17	4	41.4	3	80.0	2	83.14	1	15
20	129.74	5	10.13	4	41.5	3	66.6	2	86.44	1	15
2	104.76	5	10.13	4	53.3	2	88.0	0	93.57	1	12
17	101.28	...	10.11	...	61.1	...	85.2	...	92.00	...	.....
15	102.19	...	13.16	...	96.1	...	100.0	...	6.87	...	.....
8	148.69	...	10.28	...	41.9	...	88.0	...	79 82	...	.....
5	159.78	...	10.29	...	53.0	...	92.0	...	85.11	...	.....
4	161.42	...	10.27	...	49.2	...	96.0	...	86.82	...	.....
7	162.38	...	10.31	...	53.6	...	76.0	...	80.93	...	.....
13	168.82	...	11.29	...	77.2	...	96.3	...	43.55	...	.....
25	171.14	...	10.57	...	67.6	...	92.5	...	53.25	...	.....
19	171.84	...	10.29	...	52.8	...	88.8	...	88.24	...	.....
28	176.95	...	10.86	...	66.6	...	59.9	...	39.24	...	.....
23	183.90	...	10.36	...	44.0	...	88.8	...	81.81	...	.....
30	221.45	...	11.58	...	85.2	...	92.5	...	28.60	...	.....
24	223.16	...	10.52	...	77.2	...	81.5	...	74.44	...	.....
29	233.43	...	11.58	...	50.4	...	85.2	...	30.29	...	.....
14	250.57	...	11.92	...	88.4	...	92.5	...	27.05	...	.....
26	253.24	...	10.96	...	71.3	...	74.1	...	51.21	...	.....
12	281.86	...	11.15	...	75.6	...	88.8	...	48.88	...	.....
11	284.04	...	10.82	...	64.6	...	96.3	...	65.63	...	.....
10	336.51	...	10.96	...	61.6	...	80.8	...	68.07	...	.....
27	425.57	...	11.75	...	71.6	...	81.5	...	50.11	...	.....

* From Table XVII, p. 58.
† From Table XIII, p. 51.
‡ From Table VI, p. 30.
§ From Table VIII, p. 34.
¶ From Table XII, p. 50.
‖ Sum of weights (wt.) given in this Table.

Conclusion

It is important to note that the 10 tests eliminated were low in each of the items considered. If the only item involved had been the size of the validity index, the result would have been practically the same. The writer's subjective estimate

of the worth of the various tests was also upheld by the results of this elimination.

As a final indication of the wisdom of this selection, the coefficient of correlation, $r = .58$, was obtained between scores on selected tests and the criterion.[23]

CONSTRUCTION OF EQUIVALENT FORMS

Equivalent forms of a test are of value for several reasons:

(1) They help to eliminate coaching.

(2) They make possible repeated testing of identical groups.

(3) They aid in the administration of the test by allowing one part, or form, to be given at a time.

(4) They assist in obtaining more reliable data on the thing measured.

The Scale of Motor Ability Tests is to be used as a single test and the construction of equivalent forms is not vital to its proper use. The forms to be suggested in this section are proposed for the assistance they may give an examiner who only is able to give a part of the tests at one time. Several forms have been referred to in various Tables presented in this chapter and elsewhere.

None of the forms presented will be of identical difficulty, but all will have approximate relative equality. When the T scale technique is used and a separate scale furnished for each form, T scores on different forms will be directly comparable even if the forms are not of equal difficulty.[24]

Forms A and B were made by arranging the 30 test elements into two sets of 15 tests, each equated on the basis of the difficulty of each test as determined by the per cent of pupils passing each test.[25] These two forms were mainly purposed to facilitate trial administrations of the tests. *Form A* contains tests numbered 1 to 15, inclusive, arranged in order

[23] Judgment ratings for the Baldwin group.

[24] W. A. McCall, "How to Measure in Education," pp. 305-306.

[25] Based upon the Baldwin group of 451 pupils.

of difficulty. *Form B* contains tests numbered 16 to 30, inclusive, arranged in order of difficulty.[26] The self-correlation coefficient, r = .71, has been reported on page 35. Should any one wish to use all 30 of the Motor Ability Tests, they should use Forms A and B.

Other investigators or teachers who may wish to use the 30 tests of Forms A and B but who wish the tests to be equated on the basis of their validity as determined by their criterion scores for successes and failures, should use Forms R and S, weighting each test passed in accordance with the weight given in Table XIV, page 53.

Forms C and D may be used if it is desired to obtain a pupil's score on balance tests alone. These forms include only balance tests and are equated upon the basis of difficulty. Table XX, page 64, gives the tests included in Forms C and D.

TABLE XX

BALANCE TESTS IN FORMS C AND D

Test Number (As in Forms A and B)	Difficulty Per Cent Passed	Test Number (As in Forms A and B)	Difficulty Per Cent Passed
1	98.00	16	95.56
2	93.36	17	92.01
3	91.35	20	85.58
18	90.68	23	81.37
22	79.37	9	68.95
28	38.80	10	67.62
Sum	491.56	Sum	491.09

That the tests in the preceding Table[27] are valid tests of

[26] Described on pp. 2 ff.
[27] Table XX, p. 64.

a pupil's ability to control his balance is not fully supported by evidence in this study. They do, however, call for control of balance. The coefficient of correlation for scores on Form C plus Form D with the judgment rating criterion was

$$r = .46$$
$$P.E. = .025$$ [28]

Forms E and F may be used if it is desired to largely reduce the number of tests involving balance control. Forms E and F were approximately equated upon the basis of difficulty.

Form E includes tests numbered [29] 4, 5, 6, 7, 11, 27, 12, 29, 15, listed in order of difficulty.

Form F includes tests numbered 19, 21, 22, 24, 25, 26, 13, 30, 14, listed in order of difficulty.

The coefficient of correlation for the sum of scores on Forms E and F with the judgment rating criterion was

$$r = .57$$
$$P.E. = .021$$

Forms M and N include those tests selected, as shown in Table XIX, page 62, to comprise the final Scale of Motor Ability Tests.

Form M contains tests numbered 17, 7, 8, 24, 5, 11, 12, 13, 29, 15. They are arranged according to difficulty in the order in which they should be given.

Form N contains tests numbered 19, 4, 23, 10, 25, 26, 27, 28, 30, 14, arranged according to their difficulty, and in the order in which they should be given.[30]

The coefficient of correlation for Forms M and N with the judgment rating criterion was

$$r = .58$$
$$P.E. = .021$$

---

[28] For this correlation, judgments were arranged in five categories.

[29] These numbers correspond to the description of the tests for Forms A and B in Chapter I, pp. 2 ff.

[30] These numbers correspond to the numbers on Forms A and B described on pp. 2 ff. For a description of these tests and their numbers as they appear in the Scale, see Chapter VIII, pp. 105 ff.

The self-correlation coefficient for Form M with Form N was

$$r = .61$$
$$P.E. = .020$$

The final Scale of Motor Ability Tests gives scale scores for the sum of tests passed for the 20 tests known as Forms M and N.

Conclusion

The construction of a scale of tests does not necessitate the preparation of equivalent forms. Such forms are, however, an advantage.

While the various sets of forms enumerated in this section are not of complete respective equality, they are nearly enough equivalent to satisfy purposes of administration. Scores on any two forms may be made comparable by constructing a T scale for each form and comparing T scores.

If all 30 tests are to be given, Forms A and B may be used, while if the final scale of tests is to be given, Forms M and N must be used.

SELECTION AND APPLICATION OF A SCALING TECHNIQUE

A detailed discussion of various methods of scale construction and their respective merits need not be included in this study. McCall has admirably discussed this problem.[31]

Many tests and a few scales are now used in physical education,[32] but it is a lamentable fact that practically none of these scales is statistically sound, and that many of the tests are poorly constructed. In a few localities percentile scales are in use.[33] The writer has advocated and for three years had used at Teachers College, Columbia University, the

[31] William A. McCall, "How to Experiment in Education," pp. 94-98; "How to Measure in Education," Chapter IX, pp. 249-271.

[32] Samples of so-called Physical Efficiency Tests are given in "The Proceedings of the Athletic Research Society for 1921."

[33] Notably in the California Decathlon Tests of Hetherington.

McCall T-Scale technique for scaling tests in physical education.[34]

The T-scale technique has been used in this study because of its marked advantages and because its use will facilitate the inclusion of such scales as the Scale of Motor Ability Tests here proposed in the general scheme of educational tests and measurements. The construction of the T-scale has been fully explained by McCall.[35] It is also illustrated in Table XXI, page 68.

The problem of applying the T-scale technique in scaling the Motor Ability Tests is that of constructing a single scale for children of ages 8 to 18, and a single scale for college women. It is unfortunate that sufficient data is not now available with which to construct a corresponding scale for college men.

The T scale is based upon unselected twelve-year-old children. The unit of the scale is one-tenth S.D. of twelve-year-old children. When the T scale is used for groups whose variability differs from that of unselected twelve-year-olds, the fact should be indicated.

It will be seen from the Table of Age Norms for Children for the Scale of Motor Ability Tests [36] that the widest range of ability is found among twelve-year-olds. The frequencies of this age are too limited, however, to warrant basing the T scale wholly on twelve-year-old children.

The method used here is one which is intended to average the distribution for those ages around twelve, namely, eleven, thirteen and fourteen, for which most data is available, in such a way that the frequencies of twelve and thirteen-year-old children will enter more largely into the T scale. This will approximately base the scale on the scores of twelve-year-olds. The range of the T scale is then widened by correcting the T

---

[34] D. K. Brace, "A Method of Constructing Athletic Scoring Tables," *The American Physical Education Review,* April, 1924.

[35] W. A. McCall, "How to Measure in Education," pp. 272 ff.

[36] Appendix I, pp. 129 ff.

scores of fourteen and fifteen-year-olds. Since the lower ages for which data is available represent a very highly selected group, the scale has not been corrected for ages below eleven. Its range will permit, however, the testing of children of much lower age.

Although the resulting T scale may have very slight statistical inaccuracies, it has the tremendous advantage of providing a single scale for all ages from 8 to 18 years.

Construction of a T scale is illustrated in Table XXI, page 68, by the T scale for eleven, twelve and thirteen-year-old children.

TABLE XXI

COMPUTATION OF T SCALE FOR 11–12–13 YEAR OLDS*

Score	Frequency	1 Number Exceeding Each Score	2 Half Those Making Each Score	3 Sum Column 1 + 2	4 Percent Column 3 ÷ N	5 T Scale
0	1	339	.5	339.5	99.8	20
1	0	339	0	339.0	99.7	23
2	1	338	.5	338.0	99.7	23
3	3	335	1.5	336.5	98.9	27
4	4	331	2.0	333.0	97.9	30
5	9	322	4.5	326.5	96.1	32
6	14	308	7.0	315.0	92.6	36
7	20	288	10.0	298.0	87.7	38
8	23	265	11.5	276.5	81.3	41
9	27	238	13.5	251.5	74.0	44
10	31	207	15.5	222.5	65.4	46
11	50	157	25.0	182.0	53.5	49
12	36	121	18.0	139.0	40.8	53
13	32	89	16.0	95.0	27.9	56
14	38	51	19.0	70.0	20.6	58
15	14	37	7.0	44.0	12.9	61
16	20	17	10.0	27.0	7.94	64
17	12	5	6.0	11.0	2.97	69
18	3	2	1.5	3.5	1.03	73
19	1	1	.5	1.5	.442	76
20	1	0	.5	.5	.147	80

N = 340.
* Baldwin Group.

The steps in the construction of the scale for all ages are as follows:

(1) Construct a frequency distribution for each age, 8 to 18.[37]

(2) Sum the frequencies of 11, 12 and 13-year-old children for scores, or tests passed.

(3) Construct a T scale for 11, 12 and 13-year olds.

(4) In a similar way, construct a T scale for 12, 13 and 14-year-olds.

(5) Average the T scales for the two groups in Steps 3 and 4 above.

(6) Construct a T scale for 15 and 16-year-olds.

(7) Determine the difference between T scores obtained in Step 5 and T scores obtained in Step 6 for each score (number of tests passed) and find the average difference.

(8) Correct the T scale for 15 and 16-year-olds (Step 6) by adding the average difference obtained in Step 7 to corresponding T scores for 15 and 16-year-olds.

(9) Average the 11-14-year-old T scores (Step 5) with the corrected 15-16-year-old T scale (Step 8), thus obtaining the final T scale for all ages.[38]

A separate scale is furnished for women. The Table giving Age Norms for College Women[39] shows that scores for college women tended toward a uniform distribution with little change by ages. It was decided that one T scale based upon lumped scores for ages eighteen to twenty-two inclusive would be satisfactory for all ages. Such a scale was prepared without correction for other ages. It is included in the Test Booklet.[40]

---

[37] See "Age Norms," Appendix I, p. 129.

[38] Table XXII, p. 70, shows the steps enumerated above.

[39] Table XXIX, p. 127.

[40] Test Booklet, Chapter VIII, p. 100 ff.

TABLE XXII

Extending the Range of the T Scale

Step** / Score†	3 T Scale for Ages 11-12-13	4 T Scale for Ages 12-13-14	5 Average T Scores Ages 11-13 with 12-14	6 T Scale for Ages 15-16	7 Differences Step 5 with Step 6	8 Corrected T Scale for Ages 15-16	9 Final T Scale Average Step 5 with Step 8
0	20	20	20	12 *	5 ‡	17	19
1	23	22	23	15 *	5	20	22
2	23	22	23	18 *	5	23	23
3	27	25	26	21 *	5	26	26
4	30	28	29	24	5	29	29
5	32	32	32	29	3	34	33
6	36	34	35	30	5	35	35
7	38	37	38	32	6	37	38
8	41	39	40	35	5	40	40
9	44	42	43	37	6	43	43
10	46	44	45	40	5	45	45
11	49	47	48	43	5	48	48
12	53	51	52	46	6	51	52
13	56	54	55	50	5	55	55
14	58	57	58	53	5	58	58
15	61	60	61	56	5	61	61
16	64	63	64	60	4	65	64
17	69	67	68	63	5	68	68
18	73	71	72	66	6	71	72
19	76	77	76	70	6	75	76
20	80	80	80	76	4	81	80
N	340	402		186			

* Estimated T Scores.

** Step numbers correspond with steps enumerated on p. 69.

† Motor Ability Scale Tests passed. For a description of these tests see Chapter VIII, pp. 100 ff.

‡ The average = 5, added to scale scores in Step 6 to obtain scores in Step 8.

SUMMARY

It was the purpose of this chapter to present the data and methods used in scaling the tests. The problems presented for solution were (1) Determination of the optimum scoring

key, (2) Selection of tests to be included in the Scale of Motor Ability Tests, (3) Formation of equivalent forms, and (4) Selection and application of a scaling technique.

Several methods for selecting an optimum scoring key were tried. When the 30 tests of Forms A and B were scored without weighting, the correlation with the criterion (judgment ratings) was .57. When these tests were weighted on the basis of difficulty, the correlation was .55. When the number of tests was reduced to 20 and correlated with the same criterion, the correlation obtained was .58. Weighting the tests on the basis of the difference between criterion scores for tests passed and tests failed produced a correlation of .59. An empirical method of determining the validity index of each test was used and tests were weighted on this basis. Determination of the validity index involves the use of the mean criterion scores for tests passed and failed. The correlation obtained with the criterion in this case was .60.

It would appear that a scheme of weighting tests which considers both the mean criterion score of tests passed and of tests failed, and also the corresponding frequencies, produces the best results.

It was found, however, that certain tests could be eliminated on the basis of the size of their validity indices, and scored without weights, and a correlation with the criterion could still be obtained which was practically as high as that obtained when the optimum system of weighting was used. Because of its practical advantages and in light of its validity, the method of scoring tests finally selected involves only the simple total of tests passed.

In order to reduce the number of tests and still retain necessary qualifications 10 tests were eliminated, leaving 20 remaining as described in the Test Booklet.[41] The considerations used in eliminating tests included data on: (1) The validity of each test element; (2) The reliability of pupils' scores on each test element; (3) The per cent of constancy

[41] Chapter VIII, p. 100.

between successive trials; (4) The size of the mean criterion score for successes on each test, and (5) The difficulty of each test. Had the validity index of each test been the only factor considered, results would have been the same.

Equivalent forms are included in this chapter. Four principal sets of approximate equality are suggested. Forms A and B include all tests used in this study. Forms C and D include only tests involving mainly balance control. Forms E and F include tests other than balance tests. Forms M and N include the 20 tests of the Scale of Motor Ability Tests.

The T-scale technique was used for scaling the tests. A single scale was constructed which centered about the scores of twelve and thirteen-year-old students. The range of this scale was widened to allow its use with older students as well. The distribution of scores for college women permitted the construction of a single scale based upon ages eighteen to twenty-two.

CHAPTER VI

SOME APPLICATIONS OF THE SCALE OF MOTOR ABILITY TESTS

It is the purpose of this chapter to present data concerning some applications of the tests, and to indicate other uses.

Data for the following experimental applications will be given:

(1) Correlations: Motor Ability scores with scores on certain physical education achievement tests.

(2) Correlations: Motor Ability scores with age.

(3) Correlations: Motor Ability scores with weight.

(4) Correlations: Motor Ability scores with intelligence.

Suggested Applications

(5) Use of the Motor Ability (M.A.) score as a basis for measuring accomplishment in physical education.

(6) Use of Motor Ability score for purposes of classification for physical education class work.

(7) Possible use for diagnosing special performance disabilities in physical education.

(8) As a basis for assisting in equating groups for experimental studies in physical education.

(9) As a basis for experimental studies relating to the effect or value of various types of physical education activities in improving Motor Ability.

(10) As a criterion for evaluating other tests of Motor Ability which may be proposed by other investigations.

CORRELATIONS: SCORES ON CERTAIN ATHLETIC GAME SKILLS AND ATHLETIC EVENTS WITH THE MOTOR ABILITY TEST

It is the special purpose of this section to show some correlation coefficients for the Motor Ability Tests with scores

on certain athletic game skills and athletic events, which the writer has been developing and using for the past three years.

ACHIEVEMENT TESTS IN BASKETBALL SKILLS[1]

(1) Shooting Baskets

The subject stands behind a line formed by extending the foul throw line to the left to intersect the arc of the free throw line on the left side, facing the basket.[2] Ten tries are allowed and the "push shot" must be used.[3] Two points are scored if the ball goes through the basket. One point is scored if the ball does not go through the basket but strikes the rim of the basket before hitting the backboard. The highest possible score is twenty points.[4]

(2) Dribble and Shoot

The subject must stand behind the free throw line and execute one legal dribble around an Indian club standing on the point where the free throw lane line becomes the arc of a circle, on the right hand side. The ball must be caught on the bounce and tossed at the backboard in a continuous motion. Two points are scored if the ball goes through the basket and one point is scored if it does not go through the basket providing it first hits the back board. Ten tries are allowed and a possible score of 20 points may be made.

(3) Single Overhand Throw at a Target

The subject stands behind a line 30 feet from the target and throws the ball using a single overhand throw. The target is a canvas sheet hung so that the bottom edge just

---

[1] These tests are described in detail by the writer in the *American Physical Education Review,* April, 1925. Certain changes have, however, been made since that time, and are included in the above description.

[2] The lines referred to are painted on the floor of a basketball court.

[3] The ball leaves the hands at chest height or above.

[4] A T-scale is provided for each test of every sort used, and points are turned into T scale scores for each test. For a discussion of the construction of the T-scale, see Wm. A. McCall, "How to Measure in Education," pp. 272 ff, or D. K. Brace, "A Method for Constructing Athletic Scoring Tables, *Physical Education Review,* April, 1924.

touches the floor. The target is a double rectangle painted on the canvas. The outer rectangle is 48 inches wide and 60 inches high. The inner rectangle is 12 inches smaller each way. The bottom line of the outer rectangle is 14 inches above the floor.

Five tries are allowed. Two points are scored for each throw striking the inner area or its boundary lines, and one point for each throw striking the outer area or its outer boundary lines. A possible score of 10 points may be made.

(4) Push Pass at a Target [5]

The subject stands behind a line 25 feet from the target.[6] The target is the same as used in the Single Over Hand Throw. Five tries are allowed and points are counted as in Single Over Hand Throw. A possible 10 points may be made.

(5) Speed Pass

The subject stands behind a line 8 feet from a brick, or other solid, wall and using the push pass bounces the ball against the wall ten times, catching and passing as rapidly as possible. Preliminary trials are allowed.

The ten throws are timed with a stop watch from the instant that the first throw strikes the wall until the tenth throw strikes the wall.

(6) Jump and Reach [7]

The subject stands with the right side against a wall and holds a short piece (half length) of chalk in the right hand. The right arm is then reached up as high as possible and a mark is made on the wall at the top of the reach. The subject then swings his arms, jumps up, and makes a second

---

[5] In the push pass, the ball is thrown by extending both arms at or above waist height.

[6] For men or high school boys, this distance should be 30 feet.

[7] This test is somewhat like a "Physical Test of a Man," proposed by Sargent, *School and Society*, p. 128, January 29, 1921.

mark on the wall as high up as possible. The score consists of the number of inches, read to the nearest half inch, between the first mark and the second mark.

The total score made by an individual on the six tests is the mean of the T scores made on each of the six tests.

Table XXIII, page 76, gives the T scales for the six tests.

TABLE XXIII

T Scales for Six Basketball Achievement Tests (1)

Single Overhand Throw (2)		Push Pass (3)		Shooting Baskets (4)		Dribble and Shoot (5)		Speed Pass (6)		Jump and Reach (7)	
Scores	T Scale Score	Scores	T Scale Score	Scores	T Scale Score	Scores	T Scale Score	Scores 2nds 5ths	T Scale Score	Scores Inches and Halves	T Scale Score
0	27	0	*17	0	26	0	22	20.0-20.3	24	9.0-9.49	25
1	33	1	*21	1	29	1	25	18.4-19.1	27	9.5-9.99	29
2	37	2	24	2	31	2	25	18.1-18.3	29	10.0-10.49	32
3	41	3	27	3	31	3	26	17.3-18.0	30	10.5-10.9	35
4	44	4	30	4	34	4	29	17.0-17.2	30	11.0-11.4	38
5	47	5	34	5	36	5	30	16.2-16.4	30	11.5-11.9	40
6	51	6	37	6	39	6	32	15.4-16.1	31	12.0-12.4	42
7	55	7	40	7	42	7	34	15.1-15.3	32	12.5-12.9	44
8	59	8	45	8	45	8	36	14.3-15.0	34	13.0-13.4	47
9	65	9	51	9	49	9	38	14.0-14.2	37	13.5-13.9	51
10	72	10	61	10	51	10	41	13.2-13.4	38	14.0-14.4	54
				11	54	11	44	12.4-13.1	40	14.5-14.9	55
				12	57	12	48	12.1-12.3	42	15.0-15.4	58
				13	60	13	52	11.3-12.0	45	15.5-15.9	60
				14	62	14	56	11.0-11.2	50	16.0-16.4	62
				15	66	15	58	10.2-10.4	53	16.5-16.9	63
				16	70	16	62	9.4-10.1	58	17.0-17.4	65
				17	74	17	66	9.4- 9.3	66	17.5-17.9	67
				*18	80	18	70	8.3- 9.0	73	18.0-18.4	70
				*19	86	19	73	8.0- 8.2	78	18.5-18.9	75
				*20	92	20	78	*7.2- 7.4	84	19.0-19.4	80

* T Scores are estimated.

(1) Since the T Scale is constructed for unselected twelve-year olds, it is not fully reliable with groups having different variabilities. Its use, as illustrated here, has been found to be roughly satisfactory.

(2) Based upon 242 cases of college women majoring in physical education.

(3) Based upon 88 cases. From the standpoint of test construction, this test is not fully satisfactory.

(4) Based upon 178 cases.

(5) Based upon 194 cases.

(6) Based upon 224 cases.

(7) Based upon 94 cases.

That the basketball achievement tests really measure skill in basketball technique is felt subjectively by those working with the tests and is shown statistically by the correlation coefficient for the test scores and judgment ratings of the students' ability in actual basketball play.[8]

This coefficient, for a group of 90 college women, was

$$r = .55$$
$$P.E. = .053$$

The correlation coefficient for the Motor Ability Test scores with the basketball achievement scores for a group of 90 college women was

$$r = .24$$
$$P.E. = .067$$

Conclusion

The data just presented dealt with the correlation of scores on the Motor Ability Test with scores on basketball technique. There is seen to be a small positive correlation between the two sets of scores. One would not expect such a correlation to be high for the reason that the basketball achievement tests measure learning quite as much, if not more than natural ability. Many of the students in this group had had no previous experience in basketball, except intermittent play, in the two months preceding the testing. Many others had had much previous experience as players on college teams.

Although the data is not available, a higher correlation might well be expected from a group more homogeneous as to previous experience. There is also evidence to the effect that game skill tests will not correlate highly with general athletic ability.[9] In Table III, page 15, basketball was ranked

[8] These judgment ratings were on a scale of ten and were based upon a judgment of the girls' ability to play basketball. The rating, therefore, considers not only technique, but power and skill in playing methods. The judgments were made by Miss Marjorie Hillas, Coach of Basketball and Instructor in Physical Education, Teachers College, Columbia University, New York.

[9] J. F. Rogers, *op. cit.*, p. 23, footnote 21.

fifth as an activity in which skill might be regarded as a criterion of motor ability.

ACHIEVEMENT TESTS IN INDOOR BASEBALL

Four tests of indoor baseball technique have been used. A brief description of the tests follows:[10]

(1) Batting

The regulation indoor baseball bat and a 14-inch indoor baseball is used. The subject stands at a home plate which is opposite the center of a thirty-foot line, twenty feet away. The line is drawn at right angles to the imaginary line from home to second base and extends for fifteen feet on each side of the home to second base line. The ball is tossed from the twenty-foot line with uniform speed by one person. The batter places the balls alternately on the right and left side of the pitcher.

Six tries in all, three to each side, are allowed.[11] Each batted ball crossing the scoring line, on the intended side, in the air counts two points and one point if it crosses the line on the bounce or roll. A possible twelve points may be made.

(2) Speed Throw

The subject stands behind a line eight feet from a brick or other solid wall, and throws an indoor baseball against the wall, catching it on the rebound. Ten throws are taken in succession and as rapidly as possible. The score is the elapsed time from the instant the first throw strikes the wall to the instant the tenth throw strikes the wall, taken with a stop watch.

---

[10] For a description of similar tests and scale scores for a group of college men, see C. D. Wardlaw "Fundamentals of Baseball," Chapter VII, "Testing the Mechanics of Baseball," by D. K. Brace. Scribners, N. Y., 1924.

[11] A better test would be allow five trials for each side.

(3) Under Arm Pitch at a Target

The subject stands behind a line thirty feet from the target and pitches the ball in the regulation underarm style for indoor baseball. The target is a canvas sheet hung just above the floor. On it are painted in black two rectangles. The outer rectangle is 30 inches wide by 48 inches high and its lower line is 14 inches above the floor. The inner rectangle is six inches smaller each way than the outer rectangle.

Ten tries are allowed. Two points are scored for each ball striking the center area or its boundary lines, and one point is given for each ball striking in the outer area or on its outer boundary lines.

(4) Overhand Throw at a Target

The subject throws from behind a line forty feet from a target hanging just to clear the floor. The target is painted on a canvas sheet and consists of two rectangles. The outer rectangle is 44 inches wide and 60 inches high. The inner rectangle is 12 inches smaller each way. The lower line of the outer rectangle is 6 inches above the floor.

Ten tries are allowed. Two points are counted for hits in the inner rectangle or on its boundary lines, and one point is counted for hits in the outer rectangle, or on its outer boundary line.

Point scores on these tests as for all other skill tests are converted into T-scale scores. The individual's total score is the mean of the T scores on each separate test.

Comparative data for the Motor Ability Test and the baseball achievement tests are available for a group of 58 college women.[12] Table XXIV, page 80, gives the T scales for this group.[13]

---

[12] Junior women majoring in Physical Education at Teachers College.

[13] For similar tests with a group of men, see C. D. Wardlaw, *op. cit.*, Chapter VII by D. K. Brace.

TABLE XXIV

T Scales for Baseball Achievement Tests

Underhand Pitch		Overhand Throw		Rapid Throw		Batting	
Scores (Points)	T Scale Scores	Scores (Points)	T Scale Scores	Scores (Seconds and Fifths)	T Scale Scores	Scores (Points)	T Scale Scores
0	*17	0	*16	14.3-14.4	*22	0	22
1	*21	1	*20	14.1-14.2	*25	1	26
2	26	2	26	13.4-14.0	28	2	33
3	31	3	32	13.2-13.3	*30	3	37
4	35	4	36	13.0-13.1	34	4	42
5	38	5	40	12.3-12.4	36	5	46
6	41	6	43	12.1-12.2	*38	6	50
7	44	7	46	11.4-12.0	40	7	56
8	47	8	50	11.2-11.3	44	8	60
9	51	9	54	11.0-11.1	47	9	65
10	54	10	56	10.3-10.4	49	10	71
11	57	11	57	10.1-10.2	51		
12	62	12	61	9.4-10.0	57		
13	65	13	64	9.2- 9.3	64		
14	66	14	67	9.0- 9.1	69		
15	67	15	71	8.3- 8.4	*74		
16	70	16	*75	8.1- 8.2	*79		
17	74	17	*79	7.4- 8.0	*85		
18	*78	18	*83				
19	*82	19	*88				
20	*86	20	*92				

* Estimated.

The total score of an individual on the baseball achievement tests is the mean of the scale scores on each separate test. It is not assumed that these tests sample or measure all of the factors of baseball. They do, however, sample certain of the basic techniques of the game.

The correlation coefficient for scores on the Motor Ability Test with scores on the baseball tests for the group of 58 college women was

$$r = .168$$
$$P.E. = .088$$
$$P.I. = .014$$

This coefficient shows a slight tendency of Motor Ability scores to vary with scores on the baseball tests.

Conclusion

The coefficient of correlation obtained for Motor Ability Tests with baseball achievement tests for a group of college women was r = .168, with a relatively high P.E. These results must be regarded as inconclusive. Further testing with a larger group would, however, probably produce a correlation not greatly higher.

CORRELATION: ACHIEVEMENT TESTS IN INDOOR SOCCER FOOTBALL WITH THE MOTOR ABILITY TESTS

The fundamental techniques involved in soccer football may be taught in a gymnasium to physical education classes. Although the situation differs widely from that of the outdoor field, nevertheless the ball must be controlled in the same manner and manipulated in much the same manner, indoors as outdoors.

Four achievement tests, sampling the skills involved in the game, have been used for three years by the writer with college men and women. Changes and improvements in the tests have been made from time to time.[14] [15]

In brief, the tests are given as follows:

(1) Dribbling

Regulation soccer balls, tightly inflated, are used in all tests.[16] Four Indian clubs are set up on the floor ten feet apart. Three feet from the first club, a short starting line is drawn.

---

[14] A description of these tests as given in 1925 will be found in the January, 1925 issue of *The Discobolus,* published by the Physical Education Association of Teachers College.

[15] For a description of the game of soccer as played by women, see Frost and Cubberly, "Field Hockey and Soccer for Women."

[16] Volley balls may be substituted for women inexperienced in "heading" a soccer ball.

The subject stands with the ball at his feet on the starting line and dribbles the ball in figure-eight fashion in and out around the four clubs, returning again to the starting line.

The score is the time, taken with a stop watch, from the instant that the ball is first moved to the instant that the ball crosses the finish line again.

(2) Heading

The subject heads a ball, tossed head high to him, through a goal. The goal consists of jumping standards [17] placed on a line so that there is a central goal eight feet wide and a smaller area each side, two feet wide, as shown in the accompanying diagram.[18]

The subject stands ten feet from the goal and opposite its center, with his right side toward the goal. An assistant stands facing the subject, ten feet from him, and tosses a soccer ball so that it will come down approximately on the head of the subject. The subject is required to "head" the ball through the goals with the side of his head.

A ball going through the center goal counts two points, and one going through either of the side goals counts one point. A ball striking any post and not going through any goal counts one point. Five tries are allowed, thus making the highest score ten points.[19]

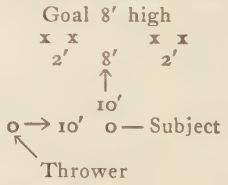
(3) Kicking Goals

A goal exactly similar to the one used for "Heading" is set up. Facing this goal and thirty feet from it a line is drawn. The subject starts behind this line, dribbles the ball

---

[17] Vertical lines may be drawn on a wall in place of the jumping standards.

[18] Goal for "heading."



[19] The test is improved if ten tries are allowed and if more points are awarded the balls going through the goal on the fly than on the roll or bounce.

twice and then kicks it toward the goal. The kick must be made from behind the line and while the ball is in motion.

Five tries are allowed. Two points are scored for each ball kicked through the center goal and one point for each ball going through the outer areas. A ball striking a post and not going in, counts one point. The ball must not be over eight feet from the floor on going through the goal.

(4) Throw In

The subject executes a regulation "throw in" [20] from a line at a target on the floor. The target is made by drawing a five by eight foot rectangle on the floor with a circle of three feet diameter on its center. The target is so placed that the long axis is at right angles to a throwing line thirty feet away from the center of the target.

Five throws are allowed. Two points are counted for each ball striking the central part of the target and one point for each ball striking the outer part of the target.

Scores on the soccer achievement tests and on the Motor Ability Test are available for a group of 61 college women in physical education.

Table XXV, page 84, gives the T scales on the soccer achievement tests for the group mentioned.

The correlation coefficient for scores on the Motor Ability Test with the soccer achievement tests was [21]

$$r = .31$$
$$P.E. = .078$$
$$P.I. = .050$$

Conclusion

The coefficient of correlation of $r = .31$ appears to show that there is somewhat marked evidence that abilities as

[20] In throwing in the ball from the side lines the player must stand on the line and throw the ball with both hands from behind and over the head.

[21] For a group of 61 college women.

measured by the Motor Ability Tests and by the soccer achievement tests tend to accompany each other. Students above the average in the soccer tests will tend to be above

TABLE XXV

T Scales for Soccer Achievement Tests

Dribbling		Heading		Kicking Goals		Throw-In	
Scores (Seconds and Fifths)	T Scale Scores	Scores Points	T Scale Scores	Scores Points	T Scale Scores	Scores Points	T Scale Scores
31 & above	20	0	28	0	22	0	
30-30.4	24	1	32	1	27	1	23
29-29.4	25	2	37	2	32	2	27
28-28.4	26	3	40	3	37	3	30
27-27.4	28	4	43	4	42	4	34
26-26.4	30	5	47	5	46	5	38
25-25.4	32	6	50	6	50	6	43
24-24.4	33	7	54	7	55	7	48
23-23.4	35	8	57	8	61	8	54
22-22.4	36	9	62	9	66	9	60
21-21.4	38	10	68	10	72	10	68
20-20.4	40						
19-19.4	42						
18-18.4	44						
17-17.4	47						
16-16.4	50						
15-15.4	54						
14-14.4	59						
13-13.4	64						
12-12.4	70						
11-11.4	74						
10-10.4	79						

the average in the Motor Ability Tests.[22] However, this tendency is not highly marked.

Were the same tests (soccer) given to pupils of equal initial experience, a higher coefficient than that obtained might well be expected.

[22] H. O. Rugg, *op. cit.*, pp. 256-257.

CORRELATION: SCORES ON THE MOTOR ABILITY TEST WITH SCORES ON INDOOR TRACK AND FIELD EVENTS [23]

Scores on 17 events of the "indoor track meet" sort were correlated with Motor Ability Test scores for a group of 64 college women.[24] The events were as follows: Dash, rope vault, high jump, hop-step-and-jump, shot put, hurdle races (two), broad jump, potato race, dips, leaning pull-ups, chinning, medicine ball pitch, standing leap, standing high jump, two lap run, jump and reach.

The zone scoring plan was used for all events except the two lap run and the potato race, which were timed and scaled. In the zone system of scoring, points are given for achievement by areas on a five point scale. The exact distance jumped, for example, was not measured, but points were given according to the area into which the contestant jumped, the areas having been previously measured and marked on the mat. This method of scoring distinguishes degrees of ability rather well when many events are used.

All point scores were totaled and averaged for each individual. A T scale was constructed for these average scores and each individual was given a T score corresponding to her average.

The coefficient of correlation for the T scores on indoor "track events" with Motor Ability Test scores was [25]

$$r = .56$$
$$P.E. = .057$$
$$P.I. = .172$$

---

[23] Except in the case of running, these events were scored by what might be called the zone system, in which points are awarded for scores within certain rather wide limits. The method is advantageous for rapid scoring of class competitions. While it does not give the accurate scoring obtained by finer measurements, it does, over a range of activities, differentiate contestants as to ability.

[24] Teachers College Juniors in physical education.

[25] Scores were on the 30 tests of Forms A and B, p. 2.

Conclusion

The method of scoring individuals on the "track and field" events was by the zone system. Scores made in the events were changed into T scale scores.

The coefficient of correlation, $r = .56$, for the indoor "track and field" events with the Motor Ability Tests indicates, for this group, a marked correspondence between the abilities as measured by these sets of tests.

While the data presented here involves the use of rather small groups, in some cases, the indications are very small that high correlation would be obtained for the Motor Ability Tests with athletic game skill (achievement) tests. If the factor of previous practice could be eliminated, higher correlations than those reported should result.

It may be inferred, therefore, that the practice of calling achievement tests, tests of motor ability is not based upon sound evidence. Track and field events come much nearer being tests of general motor ability, but even in such events variations in amounts of practice are usually encountered.

CORRELATIONS: MOTOR ABILITY TESTS WITH AGE

One of the interesting problems in the study of motor ability is the relation between growth in motor ability and chronological age. Intelligence, as measured by mental age, increases on the average at least until the age of thirteen or fourteen. Does motor ability grow in a similar way, or does growth continue indefinitely or for a longer period?

To the extent that the scale of tests proposed in this study measures pure motor ability, an indication of growth in this trait may be obtained from the norms in Appendix I, page 129.

The coefficient of correlation for age with motor ability [26]

[26] The 30 tests of Forms A and B were used.

was obtained for two groups. For the Baldwin group of 451 pupils, boys and girls, this coefficient was

$$r = .18$$
$$P.E. = .031$$
$$P.I. = .017$$

For college women, the coefficient of correlation for age with motor ability [27] was

$$r = .217$$
$$P.E. = .033$$
$$P.I. = .034$$

Sufficient data for college men is not available to include here a correlation of motor ability with age for men.

Conclusion

The correlation between motor ability, as shown by the tests treated in this volume, and age, is low. Success on the Motor Ability Tests is apparently not primarily dependent upon age.

CORRELATIONS: MOTOR ABILITY TESTS WITH WEIGHT

Tests which are true measures of motor ability should not be greatly influenced by weight. It appears logical, however, that extreme overweight will be a handicap in any sort of physical activity which demands handling the body with ease and control.[28]

The coefficient of correlation for Motor Ability Tests with weight [29] for the Baldwin group was

$$r = .045$$
$$P.E. = .034$$
$$P.I. = .00$$

---

[27] For 382 women, ages 13 to 35, measured by the tests of Forms A and B.

[28] As regards these Motor Ability Tests, no data in support of this position is available. F. R. Rogers, *op. cit.*, discusses some of the points involved.

[29] For the Baldwin group, 451 pupils, ages 10 to 18, using the 30 tests of Forms A and B.

For 384 college women, the coefficient of correlation for the Motor Ability Tests with weight was

$$r = -.111$$
$$P.E. = .034$$
$$P.I. = .001$$

Conclusion

There is practically no correlation between weight and motor ability as shown by these tests.

In the case of 384 college women, there is a very small negative correlation showing a very slight tendency on the average for heavier women to make poorer motor ability scores. The coefficients of correlation reported are so small that one would be safe in assuming a correlation of practically zero.

CORRELATIONS: MOTOR ABILITY TESTS WITH INTELLIGENCE

Data is available for the correlation of motor ability with intelligence, where motor ability is measured by the Motor Ability Tests and intelligence is measured by the Thorndike Psychological Examination for High School Graduates.[30]

The subjects in this data are 109 college women.[31] The coefficient of correlation obtained was

$$r = -.159$$
$$P.E. = .063$$
$$P.I. = .012$$

As a partial check on the above data, the coefficient of correlation for intelligence (tested as above) with scores on

[30] Published by Teachers College, Columbia University, New York City.

[31] Teachers College students majoring in physical education, 1925-1926. The tests used are those of Forms A and B.

the basketball achievement tests was obtained. This coefficient, for 91 college women,[32] was

$$r = -.154$$
$$P.E. = .069$$
$$P.I. = .012$$

Conclusion

It has been proposed that probably all desirable human traits tend to correlate positively with each other. It is also usually assumed that any correlation between motor ability and intelligence would be small.

The data here presented has shown that the correlation between intelligence and motor ability, as measured by the Motor Ability Tests, and between intelligence and basketball skill, is in both cases negative and practically negligible. The size of the probable errors is not such as to indicate fully satisfactory statistical reliability.

SUGGESTED APPLICATIONS

The following possible uses of the Scale of Motor Ability Tests are suggested. No data for their support will be given.

(1) Motor Ability Scale scores may be used as a basis for measuring accomplishment in physical education.

Given a pupil's T scale score in any athletic event and his Motor Ability Scale score a motor accomplishment quotient (M A Q) may be easily determined which, by its size, will show whether or not the student is achieving the results which his motor ability indicates he should achieve.

The following formula may be used in computing the (M A Q) Motor accomplishment quotient =

$$\frac{\text{Achievement score}^{33} \times 100}{\text{Motor Ability Score}}$$

[32] Some of the same students in the previous group.

[33] Where T scores are used.

where Achievement score is a T scale score on an athletic event.

For example, John Jones makes a T scale score of 40 in the 100-yard dash. His motor ability score is 60. Substituting in the formula

$$\frac{\text{Achievement score} \times 100}{\text{Motor Ability Score}} = \text{Motor accomplishment quotient we have } \frac{40 \times 100}{60} = 66.6 = \text{his motor accomplishment}$$

quotient for the 100-yard dash. It shows that John Jones has done only 66.6 per cent as well as he should have. With continued instruction and practice this pupil should reach a scale score of 60 in this event.

Another student makes a scale score of 60 in the high jump. His motor ability score is 50. His accomplishment quotient becomes 120, thus showing that his acomplishment is 20 per cent higher than the average for one with his motor ability. A pupil who is just working to the best of his ability will have a motor accomplishment quotient (M A Q) of 100.

It is also possible to obtain the motor accomplishment quotient of a class by using the class average achievement score and the class average motor ability score in the formula for finding the M A Q.

The advantages and uses of the motor accomplishment quotient are obvious. Many teachers wish to include in their grading of pupils a grade indicating effort. Heretofore this has been based simply on the teacher's judgment of how nearly the pupil has been working to the best of his ability, or upon the achievement made. While measuring actual achievement is worth while it does not give a satisfactory measure of effort for, obviously, a student of superior natural ability can make good achievement scores with little real effort,

and may, in fact, be loafing and forming accompanying bad habits.

The use of the motor accomplishment quotient, for the first time, allows the teacher to do, in physical education, what is now being done in other school subjects; namely, to give grades which are based upon achievement in terms of ability to achieve.

(2) Students could be classified, for purposes of class organization, on the basis of their motor ability. Such classifications should be coarse, rather than fine, and a fully adequate basis of classification should involve other factors also. One such factor might well be the strength index, as developed by Rogers.[34]

(3) The diagnosis of special performance disabilities as relates to physical education offers an interesting field for research. The Motor Ability Tests might assist an investigator in such studies by giving a better picture of an individual's motor reactions.

(4) Experimental studies which desire to evaluate the activities of one program of physical education with another, and for this purpose wish to use some method of equivalent groups, are at once confronted by the difficulty of finding adequate bases upon which to equate their groups.

Equivalent groups for such studies may now be equated upon the basis of their motor ability as well as upon other, previously available bases.

If no other use than this could be made of the Motor Ability Tests developed by this study, they would still help serve a tremendous need. The determination of what activities to include in a program of physical education is of vital importance to all physical educators, who as yet have made practically no use of experimental investigations for such a purpose.

(5) The Scale of Motor Ability Tests should be of assistance to other investigators studying motor ability. It will

[34] F. R. Rogers, *op. cit.*

increase, by one, the small number of reliable criteria with which to validate proposed tests of motor ability.

It will be seen, therefore, that the Scale of Motor Ability Tests developed in this study has many practical applications for the teacher of physical education. For the research student it should also prove of assistance.

CHAPTER VII

GENERAL SUMMARY

It was the purpose of this study to develop a Scale of Motor Ability Tests. This involved two main problems:

(1) Discovery of tests of motor ability, and
(2) Scaling the Tests.

Certain guiding principles were set up to aid in selecting tests. It was desired to have tests which measured natural rather than acquired motor ability, which involved a general functioning of the whole body in a variety of activities, and which were economical of administration in point of time and equipment.

From preliminary experimentation, a group of 30 tests were selected. These resemble stunts in that each test requires a definite execution. It is either passed or failed. The final scale of tests includes 20 tests which are shown to have greatest validity and reliability.

Detailed directions for administering and scoring the scale tests are given in a Test Booklet in Chapter VIII.

The tests comprising the scale are simple and definite. Carefully prepared and illustrated instructions make giving the tests easy and accurate. Tests are scored in two ways only, either passed or failed. They may be given to four pupils at a time, as an individual test in which the examiner does the scoring; they may, with special precautions, be given to small groups in which pupils score themselves; or they may be given as a group test, in which pupils score each other, by one examiner and pupil assistants to as large a class as desired. As a group test forty minutes will give ample time to complete all tests.

The data presented in this study related almost entirely

to the group method of administration in which pupils score each other.

The subjects whose scores furnish the data upon which the scale has been developed consist of 1,555 students from the fifth grade through college. Of this number the records of 775 children, boys and girls, are given in the Table of Age Distributions. 523 records of college women are given in a separate Table.[1] No data for college men has been used and no scale for such students is offered.[2]

The subjects were distributed among three elementary schools, two junior-senior high schools and three women's colleges. The tests were given over a period of two years. Three groups were retested one year after their first examination. Eight examiners in different localities have contributed data used in this study.

The Validity of the Tests as Measures of General Motor Ability

The wisdom of selecting the type of physical activity used by this study as test material is supported by the collective judgment of 75 mature students and teachers of physical education.

The criteria upon which the tests have been validated are of three sorts:

(1) Judgment ratings as to general motor ability, given to 451 pupils by a group consisting of an instructor and ten school athletes, high school boys and girls.

(2) Scores on a variety of athletic events.

(3) A comparison of the scores of athletic team members with those of a group of 451 pupils of which they were a part.

When judgment ratings are used as a criterion of motor ability, the lowest coefficient of correlation, .47, was obtained

[1] Appendix I, p. 129.

[2] College men may be scored by using the scoring table (p. 126) for ages 8 to 18 with a fair degree of accuracy.

with test scores of girls. The highest coefficient, .68, was obtained with test scores of boys.

Various methods of weighting Motor Ability Test scores are discussed in Chapter V. The optimum method of weighting produced a coefficient of .60 for test scores with judgment ratings for boys and girls combined. The tests finally selected to comprise the scale have a correlation of .58 with the judgment rating criterion. Considering the nature of the criterion, this must be regarded as a highly satisfactory coefficient.

The second type of criterion consists of scores on ten athletic events one year, and on six athletic events the following year. Coefficients of correlation for scores on the Motor Ability Tests with scores on athletic events range from .70 to .80. The largest coefficient was obtained when the largest number of events was included in the criterion.

This would seem to indicate that a criterion consisting of scores on very many athletic events would produce a still higher correlation.

Students, 56 in number, who were members of school teams had a mean score on the Motor Ability Tests 20 per cent higher than that of the group as a whole.

Reliability and Objectivity of the Motor Ability Tests

Reliability and objectivity is enhanced by the fact that no equipment is used in the testing; that the tests are simple and are scored only success or failure; and that instructions for administering and scoring the tests are definite and readily accessible in the form of a Test Booklet, given in Chapter VIII of this volume.

As the results of statistical techniques employed with the data used, the following findings may be reported:

(1) The mean net difference between first and second trials on the tests is small, .625, for one group of college women.

(2) When a group of 27 sixth grade boys simultaneously scored the same individuals the average agreement among their scores was 86 per cent. In order not to influence this experiment, part of the instructions regarding scoring were eliminated.

(3) For an identical group of college women, the mean net difference between two sets of scores on the Motor Ability Tests separated by approximately one year was .996, where a difference of 30 was possible.

(4) The mean net difference between two forms of the tests was 1.061.

(5) Reliability is shown by self-correlations as follows:

(a) Form A with Form B, $r = .71$.

(b) Scores one year with scores one year later for college women, $r = .66$.

(c) Scores one year with scores one year later for children, $r = .71$, for one group and $r = .82$ for another group.

(d) Failures on a first trial with failures on a second trial of the same tests, $r = .90$.

(6) Objectivity is partly established by its relation to reliability as measured by coefficients of correlation of .66, .71, .73, .78 and .90 reported in Chapter V.

(7) Objectivity was further established by a coefficient of .72 for self-scored tests by fifth and sixth grade boys, with the identical pupils and tests scored by an examiner.

(8) As shown by a comparison of mean scores there was a slight tendency for pupils to score themselves more severely than did the examiner.

Techniques Used in Scaling the Tests

The problems involved in this part of the study were:

(1) The determination of an optimum scoring key.

(2) The selection of tests to be included in the Scale of Motor Ability Tests.

(3) The selection and application of a scaling technique.

(4) An incidental problem was the formation of equivalent forms of the tests.

Several methods of selecting a scoring key were tried. Methods of weighting test scores in which weights were assigned by various combinations of the mean criterion score of pupils passing each test, and of pupils failing each test, for a group of 451 pupils, increased the correlation with the criterion from .55 to .60.

An empirical method, known as the McCall-Long Method was used to determine the validity of each test. The validity index was obtained and was used in reducing the number of tests, and also as the basis of a scoring key.

Although the McCall-Long Method is without statistical measures of reliability, its use obtained the highest correlation with the judgment rating criterion of the methods tried, and may be sanctioned on this basis.

The number of test elements was reduced from 30 to the 20 comprising the single scale. This reduction retained the tests having greatest validity. It retained the more difficult ones without impairing the ability of the scale to test the lower limits of ability. It retained those tests whose mean criterion scores were highest.

It was found that the tests finally selected, known as Forms M and N, could be scored on the simple basis of tests passed and still correlate with the criterion about as well as if they were weighted in the optimum manner.

The interests of simplicity and economy of time, therefore, threw the balance in favor of computing the score as tests passed.

The scaling technique selected is that of the T scale, thus allowing the Scale of Motor Ability Tests to be expressed in scale scores comparable with scores on other educational tests.

As explained in Chapter V, a single T scale was prepared for use with boys and girls from ages 8 to 18. Another T scale, standardized on ages 18 to 22, is given for college women. It will scale, satisfactorily, the data included in this study for ages 13 to 48.

Applications

A study of some of the applications that may be made of the Scale of Motor Ability Tests led first to a comparison of motor ability scores with scores on achievement tests in basketball, indoor baseball and indoor soccer football.

Correlations obtained from this study appear to bear out the fact that tests of game skills should be rightly called achievement tests, and should not be termed motor ability tests. They measure learning much more apparently than motor ability.

Track and field events have a higher correlation with the motor ability tests than do game skills.

The motor ability scores were found to have little correlation with age, this being .18 for children, and .22 for college women. Success in the Motor Ability Tests is little dependent upon age.

Motor Ability as measured by these tests is even less dependent upon weight than it is upon age, as shown by correlation coefficients of .04 and — .11. College women are, on the average, slightly handicapped by weight.

If motor ability is a native trait, it probably develops with age, to some limit.[3] Until this limit is reached, a positive correlation with age should be expected.

Data available for a group of college women showed a very small negative correlation between motor ability as measured by the tests and mental ability as measured by the Thorndike Psychological Examination. This coefficient was — .159. Basketball achievement tests correlated no more highly with mental ability than did the motor ability tests.

[3] The age distributions in Appendix I show this to be the case.

Certain Possible Applications Which May Be Made of the Scale of Motor Ability Tests Are, Briefly, as Follows:

(1) Motor Ability Scale scores may be used as the basis for determining an accomplishment quotient for the activities of physical education.

(2) The scale of tests should form a valuable element in a scheme of classification of pupils for class work in physical education. This should be especially valuable in programs which favor the individual development of pupils.

(3) Diagnosis of special performance disabilities may be assisted by a study of individual reactions to these tests.

(4) Experimental studies in physical education which require the equating of groups of pupils have been greatly handicapped by the lack of standardized tests of general motor ability. The Scale of Tests proposed in this study would be an invaluable aid in experimental studies using equivalent groups.

(5) Finally, this Scale of Tests should stimulate other scientific efforts in the field of tests and measurements in physical education. Leads thus far established should assist other research on the nature of motor ability.

CHAPTER VIII

TEST BOOKLET

BRACE SCALE OF MOTOR ABILITY TESTS

It is the purpose of the Test Booklet to assist the reader who may wish to use these tests by collecting, in one place, all the directions for applying and scoring the Scale of Motor Ability Tests.

The tests described and illustrated in this booklet are intended to measure general motor ability. They should be given and scored exactly as explained.

DIRECTIONS FOR THE BRACE SCALE OF MOTOR ABILITY TESTS

General Directions for Applying the Tests

(1) Before giving the tests, read and study carefully the explanation of performance and scoring for each test. Study the illustrations. Be sure that you understand what constitutes failure and what success. Try each test yourself to see what it involves. Score yourself. Remember that you can demonstrate the test even though you are unable to perform it successfully.

(2) Select a place to give the tests which will allow room to space your group properly. If possible, have girls wear gymnasium suits and boys wear tennis shoes. Older girls must be in appropriate clothes.

(3) Provide a scoring blank and a pencil for each pupil taking the tests, and an ordinary watch for yourself. Use only the printed scoring blanks.

(4) Directions for Handling Pupils:

A. Directions to be followed when the tests are to be given individually and scored by an examiner.

(a) Test four pupils at a time.
(b) Have pupils stand in a line about 5 feet apart.
(c) Say to the pupils: "You are to be given a test of your ability to do certain stunts. Listen to the explanation and watch the demonstration. Do not start until I say 'Go.' No practice trials are allowed. Try each stunt slowly and carefully and do your best. Do not talk or laugh."
(d) Have each pupil properly fill in the spaces at the top of his scoring blank. Then collect these blanks and lay them on a table or on the floor in a row and in the order in which the pupils are standing.
(e) Explain and demonstrate the first test. Explain what constitutes failure and what constitutes success. Give the test and record the score on the proper blank. Repeat this process for each test.

B. Directions to be followed when pupils are to score themselves.

(a) Test children in groups of not more than 20 to 30 pupils in a group, who are in the fifth grade or above.
(b) Have each pupil fill out a score blank.
(c) Arrange the group in some form of open order so that there is at least six feet between pupils.
(d) Explain that each child is to score himself or herself on each test by marking an X for success and an O for failure.
(e) Have at least two pupil assistants. Explain to the group that the assistants will watch them and keep a check upon how they score themselves. Have the assistants walk among them and watch the scoring.

(f) Explain and demonstrate the first test and illustrate what constitutes success and what failure.

(g) Take special pains after each test to repeat the explanations of what constitutes failure. Students whose scoring is questioned should have their names noted and be retested on another day.

(h) Allow 30 minutes for the completion of all tests by this method.

C. Directions to be followed when pupils score each other.

(a) Arrange pupils in parallel rows of even number. Have approximately 5 feet between pupils in a row, and 10 to 15 feet between rows. Pass out pencils and papers from the end of each row.

(b) Have pupils fill in the spaces for data at the top of scoring blanks. Instruct them to give their weight in street clothes with outer coats removed.

(c) Have every other row of pupils face about, thus having pupils in each two rows facing together.

(d) Have pupils exchange papers.

(e) Explain: "You are to be given a test of your ability to do certain stunts. Listen to the explanation and watch the demonstration. Do not start until I say 'Go.' No practice trials are allowed. When I say 'Go,' try each stunt slowly and carefully. Do your best. Do not talk or laugh."

(f) Number the alternate lines 1 and 2. Have all pupils in lines 1 sit on the floor. Have pupils in lines 2 lay their papers and pencils on the floor and step back two steps.

(g) Explain: "Those pupils standing will take the first 10 tests. Each person sitting down will score the person whose score blank you hold. Then we will change and those sitting will take the tests. You will all have a chance at each test."

(h) Explain: "Watch the person you are scoring. Pay no attention to any one else. If the person you are scoring does the test successfully, exactly as I have explained, mark a cross (X) in the space after the number of the test. If the test is not done correctly it is failed, so mark a zero (O) in the space by the number of the test. Each test must be scored either success, with a cross (X), or failure, with a zero (O). Be fair to the person you are scoring. I will watch to see that you do it correctly."

(i) Explain and demonstrate the first test and illustrate what constitutes failure. Do it correctly and say, "That would be scored a cross (X)." Do it incorrectly and say, "That would be scored a failure, zero (O)."
Say, "Are there any questions?"
After the tests are started the examiner should discourage any questions.

(j) Use formal control while giving the tests. Allow no talking or laughing. Do not allow a pupil to watch the scoring which is done on his or her blank.

(k) It is well to have two or three pupil assistants patrol the scoring. Allow 35 to 40 minutes when the tests are pupil scored.

(l) Give the *first 10 tests* (Form M) to lines 2. Then change, have lines 2 sit and lines 1 stand. Give the *second 10 tests* (Form N) to lines 1. Change and give the second 10 tests (Form N) to lines 2. Change and give

the first 10 tests (Form M) to lines 1. This alternating reduces practice effect.

(m) Follow explanations exactly.

(5) When all the tests are completed, have the students check on the information at the top of the blank. Have papers and pencils passed to one end of each row and collected.

(6) Score the blanks. Count the total number of tests passed and enter this number in the Tests Passed space.

(7) Change the score to its corresponding Scale Score by using the Table of scale scores, Table XXVIII, p 126, provided in this booklet.[1] Enter the scale score in the Scale Score space on the scoring blank.

(8) The scale score is a student's motor ability score.

(9) Motor Ability Scale Scores may be interpreted as follows:

TABLE XXVI

APPROXIMATE MEANING OF SCALE SCORES

Scale Scores	Approximate Per Cent	Approximate Meaning *
Above 50	50	Above Average
Below 50	50	Below Average
Between 40 and 60	66	The middle two-thirds
Above 60	16	The highest one-sixth
Below 40	16	The lowest one-sixth

* In terms of all ages combined.

[1] For example: a child performs 10 tests correctly. Find the number 10 in the first column (tests passed) in Table XXVIII, p. 126, then look across and read the scale score number in the second column opposite the number 10 (tests passed). This scale score number in this case will be found to be 45. This is this pupil's score on the tests, and is his motor ability score.

ILLUSTRATIONS OF THE TESTS WITH DESCRIPTIONS AND INSTRUCTIONS ON SCORING

Form M

Test 1

Walk in a straight line, placing the heel of one foot in front of and against the toe of the other foot. Start with the left foot. Take 10 steps in all, 5 with each foot. *Eyes open.*

Failure = 1. Losing the balance and stepping out of line.
2. Not walking in a straight line.
3. Not placing heel to toe.

a *b*

a. Starting position. *b*. Correct performance.

Figure 1

Test 2

Stand Jump into the air and clap both feet together *once*, and land with the feet apart (any distance).

Failure = 1. Landing with the feet touching each other.
2. Failure to clap the feet in the air once.

a *b*

a. Starting position. *b.* During performance.

FIGURE 2

Test 3

Lie flat on the back on the floor. Fold the arms across the chest. Raise the trunk to a sitting position. Do not raise the feet above the floor, or unfold the arms.

Failure = 1. Raising the feet above the floor. (This does not include sliding the feet, which is permissible).
2. Unfolding the arms.
3. Failure to sit up.

a *b*

a. Starting position. *b*. Correct performance.

FIGURE 3

Test 4

Stand. Fold the arms behind the back. Kneel onto both knees. Get up without losing the balance or moving the feet about.

Failure — 1. Losing the balance either going down or getting up.
2. Moving the feet after standing up.
3. Unfolding the arms.

a *b*

a. Starting and Finishing position. *b.* During performance.

FIGURE 4

Test 5

Take a front leaning rest position, *i.e.*, place the hands on the floor, arms straight, extend the feet back along the floor until the body is straight (in an inclined position to the floor).

Bend the arms, touching the chest to the floor, and push up again to straight arms. Do this *3 times* in succession. Do not touch the floor with the legs or waist.

Failure = 1. Failure to push up 3 times.
2. Failure to touch the chest to the floor each time.
3. Resting the knees, thighs, or waist on the floor at any time.

a

b

a. Starting and finishing position.
b. During performance.

FIGURE 5

Test 6

Squat on the heels with feet together and knees out, and hands between the knees with fingers touching the floor. Spring up onto both heels, with legs straight and toes up, and swinging both arms out at the side level with the floor. The feet should then be about 18 inches apart. Head up. Repeat this exercise three times (in all) rhythmically.

Failure = 1. Failure to get the arms and legs in position.
2. Failure to do it three times in succession without stopping.

a *b*

a. Starting position. *b*. Performance.

FIGURE 6

Test 7

Stand with feet together. Jump into the air and make a full turn to the left, landing on the same spot. Do not lose the balance or move the feet after they strike the floor.

Failure = 1. Failure to get all the way around.
2. Moving the feet after they strike the floor.

During performance.
FIGURE 7

Test 8

Jump into the air and clap the feet together twice and land with the feet apart (any distance).

Failure = 1. Failure to clap the feet together twice.
2. Landing with the feet touching each other.

a *b*

a. Starting position. *b*. During performance.

FIGURE 8

Test 9

Stand on the right foot. Grasp the left foot behind the right knee. Bend and touch the left knee to the floor, and stand up without touching any other part of the body to the floor, or losing the balance.

Failure = 1. Touching the floor with any part of the body except the left knee.

2. Failure to touch properly and stand with right leg straight, and without losing the balance.

a *b*

a. Starting and finishing position. *b*. Performance.

FIGURE 9

Test 10

Hold the toes of either foot in the opposite hand. Jump up and jump the free foot over the foot that is held, without letting go.

Failure = 1. Letting go of the foot that is held.
2. Failure to jump through the loop made by holding the foot.

a *b*

a, Starting position. *b*. During performance.

FIGURE 10

Form N

Test 11

Jump into the air and slap both heels with the hands behind the back.

Failure = 1. Failure to touch both heels.

a *b*

a. Starting position. *b*. During performance.

Figure 11

Test 12

Stand, kick the right foot up so that the toes come at least level with the shoulders. Do not fall down on the floor.

Failure = 1. Failure to kick as high as the shoulders.
2. Falling down and touching the floor with any part of the body other than the feet.

a *b*

a. Starting position. *b*. Performance.

FIGURE 12

Test 13

Stand on the left foot. Bend forward and place both hands on the floor. Raise the right leg and stretch it back. Touch the head to the floor, and regain the standing position without losing the balance.

Failure = 1. Inability to touch the head to the floor.
2. Losing the balance and having to touch the right foot down or step about.

a. Starting (wrong foot). *b*. Correct performance.

FIGURE 13

Test 14

Stand with both feet tight together. Bend down, extend both arms down between the knees, around behind the ankles, and hold the fingers together in front of the ankles without losing the balance. Hold this position for five seconds. (Counted by scorer.)

Failure = 1. Falling over.
2. Failure to touch and hold the fingers of both hands together.
3. Failure to hold the position for five seconds.

a *b*

a. Starting position. *b*. Performance.

FIGURE 14

Test 15

Stand with both feet together. Swing the arms and jump up in the air, making a full turn to the right. Land on the same spot and do not lose the balance, that is, do not move the feet after they first strike the floor.

Failure = 1. Failure to make a full turn and land facing in the same direction as at the start.
2. Losing the balance and having to step about to keep from falling.

Starting and finishing position.
FIGURE 15

Test 16

Kneel onto both knees. Extend the toes of both feet out flat behind. Swing the arms and jump to the feet without rocking back on the toes, or losing the balance.

Failure = 1. Having the toes curled under and rocking back on them.
2. Failure to execute the jump, and stand still on both feet.

a *b*

a. Starting position. *b*. Finish.

FIGURE 16

Test 17

Fold the arms across the chest. Cross the feet and sit down cross-legged. Get up without unfolding the arms or having to move the feet about to regain the balance.

Failure = 1. Unfolding the arms.
2. Losing the balance.
3. Failure to get up.

a *b*

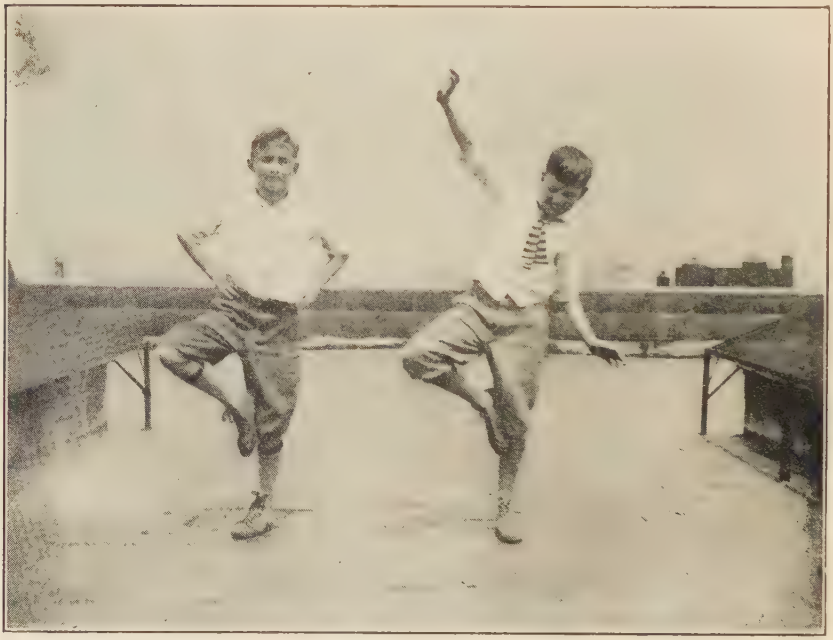
a. Starting and finishing. *b*. During performance.

FIGURE 17

Test 18

Stand on the left foot. Hold the bottom of the right foot against the inside of the left knee. Place hands on hips. *Shut both eyes,* and hold the position for ten seconds, without shifting the left foot about on the floor.

Failure = 1. Losing the balance.
2. Taking the right foot down.
3. Opening the eyes or removing the hands.



a *b*

a. Performance. *b*. Failure (hands removed from hips).

FIGURE 18

Test 19

Take a squat rest position. That is, place the hands on the floor between the knees and close to the feet. Bend the elbows slightly and place both knees *well over* the elbows. Rock forward onto the hands, raising the feet from the floor. Support the body on the hands. Hold the position for five seconds (as counted by the scorer).

Failure = 1. Failure to keep the body off the floor for five seconds.

a *b*

a. Starting position. *b*. Performance.

FIGURE 19

Test 20

Stand on the left foot with the right foot extended forward off of the floor. Sit down on the heel of the left foot, without touching the right foot or hands to the floor. Stand full up without losing the balance.

Failure = 1. Failure to sit all the way down on the left heel.
2. Touching the right foot or hands to the floor.
3. Failure to stand up with left leg straight before touching the right foot.

a *b*

a. Starting position (incorrect—standing on wrong foot). *b*. Performance.

FIGURE 20

TABLE XXVII

Scoring Blank for Brace Scale of Motor Ability Tests

Date..... Name.....

Class..... Weight.....Height.....

School..... Age ..... years..... months

Form M Form N

In Scoring X Means Success; O Means Failure

Number of Stunt	1st Trial	2nd Trial	Number of Stunt	1st Trial	2nd Trial
1			11		
2			12		
3			13		
4			14		
5			15		
6			16		
7			17		
8			18		
9			19		
10			20		

Tests Passed..... Scale Score.....

TABLE XXVIII

TABLE FOR CONVERTING TEST SCORES INTO SCALE SCORES *
For Ages 8 to 18

SCORE (Tests Passed)	SCALE SCORE
0	19
1	22
2	23
3	26
4	29
5	33
6	35
7	38
8	40
9	43
10	45
11	48
12	52
13	55
14	58
15	61
16	64
17	68
18	72
19	76
20	80

* To Use this Table: Find the sum of tests passed—(Forms M and N described in the Test Booklet). The corresponding Scale Score is the score made by the student on the Motor Ability Tests.

TABLE XXIX

Scoring Table for College Women

Motor Ability Score (Sum of Tests Passed)	Scale Score
1	12 *
2	14 *
3	17 *
4	20 *
5	24
6	30
7	34
8	38
9	41
10	45
11	48
12	52
13	56
14	60
15	63
16	67
17	71
18	75
19	78
20	80

* Estimated.

APPENDIX I

Appendix I contains two Tables of Age Distributions with Norms for certain ages.

1. Table XXX—Age Distributions for Ages 8 to 18.
2. Table XXXI—Age Distributions for College Women, Ages 13 to 48.

TABLE XXX

Age Distributions and Norms for Ages 11-16*
Boys and Girls—Ages 8-18

Tests Passed	Scale Score	Ages										
		8	9	10	11	12	13	14	15	16	17	18
0	19					1						
1	22											
2	23				1						1	
3	26					2	1					
4	29		4	1	1	2	1	1	2			
5	33	1	1	3	3	1	5	4	1	1	1	
6	35		1	7	3	4	7	1	1		2	
7	38		2	6	7	9	4	4	4		1	
8	40		4	4	9	9	5	4	4	1	1	
9	43		1	4	7	6	14	9	10		1	1
10	45		2	8	12	9	10	15	10	5	2	1
11	48		3	3	11	16	23	21	12	4		2
12	52			3	11	7	18	25	16	6	2	
13	55			1	8	8	16	16	19	6		3
14	58	1			4	17	17	13	20	6	1	1
15	61			1	5	4	5	11	12	8	5	3
16	64				1	9	10	12	6	9	2	1
17	68				2	5	5	4	9		1	1
18	72					2	1	7	4	4	3	2
19	76				1			1	2	2		2
20	80						1		1	1		
Total Pupils 775		2	18	41	86	111	143	148	133	53	23	17
Means*			7.56	8.84	10.97	11.85	11.91	12.74	13.13	14.36	12.63	14.97

* The means or age norms are most reliable for 11- to 16-year-old children.

TABLE XXXI

Age Distributions and Norms for Ages 18-22† Women—Ages 13-48

Tests Passed	Scale Score	Ages																				
		13 to 16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36 to 48
1	12*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2	14*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	17*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	20*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5	24	.	.	1	.	.	1	1	.	.	.	.	.	.	.	.	.	1	.	.	.	.
6	30	.	2	3	3	1	.	1	.	1	.	2	.	.	.	.	1	.	.	1	.	2
7	34	.	2	4	7	2	4	1	2	1	.	.	.	1	.	1	.	.	.	.	1	1
8	38	1	3	2	11	5	1	4	1	.	1	2	1	.	1	.	.	2	.	.	.	.
9	41	.	3	4	3	11	5	3	3	1	.	1	2	1	4	1	1	1	1	.	.	1
10	45	.	3	15	7	8	9	2	2	1	1	4	.	.	2	1	.	.	1	.	.	2
11	48	2	2	10	13	11	7	6	5	1	2	4	3	1	1	2	.	.	.	3	1	2
12	52	.	6	12	12	15	9	5	4	3	2	1	4	.	2	2	.	3	.	.	.	2
13	56	1	2	12	10	14	7	4	2	5	2	3	4	.	2	.	.	.	.	2	.	.
14	60	.	3	.	7	3	12	9	3	2	1	.	.	2	.	.	1	2	1	.	1	2
15	63	.	.	1	4	5	3	2	2	.	1	1	.	.	.	1	.	1	.	.	1	.
16	67	.	.	3	2	2	4	3	3	2	1	1	.	1	.	.	1	.	.	1	.	.
17	71	.	.	.	1	.	3	2	.	.	2	1	.	.	.	.	.	.	.	.	.	.
18	75	.	.	.	.	.	.	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.
19	78	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
20	80	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Total Pupils 523		4	26	67	80	77	66	45	27	18	13	20	14	6	12	8	4	10	3	7	4	12
Means†				11.33	11.45	11.81	12.00	12.70	12.39	12.17	13.50	11.50										

* Estimated.

† Means are only fairly reliable for ages 18 to 22.

APPENDIX II

APPENDIX II contains two charts referred to in the text as indicated by footnotes. The charts are scattergrams from which the indicated coefficients of correlation were computed.

CHART I

SCATTERGRAM: SCORES ON MOTOR ABILITY TESTS WITH JUDGMENT RATINGS *

Motor ability scores	Judgement ratings																				
	1	2	3	4	5	6	7										17	18	19	20	
32																			2		2
30						1	1						1	1	2	1	3	4	6	2	22
28				1	1	1	1	2	1	4	2	3	1	4	9	6	3	2	1	1	43
				2	6	4	5	6	2	10	1	6	6	10	3	8	2		1		72
		1			4	8	5	8	4	7	4	7	8	10	7	4		2	1		80
		1	1	4	9	4	6	4	2	5	2	7	7	3	5	6	1			1	68
	1		2	9	15	3	6	5	3	5	3	5	1	4	4	3	2	1	1		73
	1	5	1	7	9	4	1	4		2	1	1	1		1			1			39
	1	2	1	6	7	1			1	2		1			3			1			26
	1	1	3	2		1	1	1						1							11
	1	4	2	2							1										10
10	1	1	1		1																4
8																					0
6		1																			1
	6	16	11	33	52	27	26	30	13	35	14	30	25	33	34	28	11	11	12	4	451

* For children of ages 8-18. This chart is referred to on page 18 of the text.

CHART II

SCATTERGRAM: FIRST TRIAL FAILURES WITH FAILURES ON SECOND TRIALS OF THE SAME TESTS *

First trial failures	Second trial failures													
	0	2	4	6	8	10	12	14	16	18	20	22	24	
26													1	1
24														0
										2	2			4
							2	6		2				10
					2		3	5		1				11
					3	8	5	8	2					26
			1	4	8	10	11	5						39
		1	2	20	22	24	4							73
		4	18	22	17	6								67
		6	34	33	7									80
6	1	23	36	12										72
4	3	21	18	1										43
2	17	6												23
0	2													2
	23	61	109	92	59	48	25	24	2	5	2	0	1	451

* For children of ages 8-18. Reference to this chart is made on page 41 of the text.

BIBLIOGRAPHY

BAILOR, E M., "Content and Form in Tests of Intelligence," Teachers College, Columbia University, New York, 1924.

"Barlow's Tables," Spon and Chamberlain, New York, 1924.

BRACE, D. K., "A Method for Constructing Athletic Scoring Tables," *American Physical Education Review,* April, 1924.

——"Testing Basketball Technique," *American Physical Education Review,* April, 1924.

——"A Natural Program for Physical Education in Colleges," *American Physical Education Review,* April, 1925.

——Soccer Tests, *The Discobolus,* January, 1925.

FROST, HELEN, and CUBBERLY, HAZEL, "Field Hockey and Soccer for Women, *Scribner's,* 1923.

HETHERINGTON, C. W., "Decathlon Tests," Published by the American Physical Education Association Press.

HOLZINGER, K. J., "Statistical Tables for Students in Education and Psychology," University of Chicago Press, Chicago, Illinois.

KELLY, T. L., "Statistical Methods," Macmillan Co., 1923.

MCCALL, WM. A., "How to Measure in Education," Macmillan, 1922.

——"How to Experiment in Education," Macmillan, 1923.

"Proceedings of the Athletic Research Society," 1921.

"Report of a National Committee—Motor Ability Tests," *American Physical Education Review,* December, 1924.

ROGER, F. R., "Physical Capacity Tests in the Administration of Physical Education," Teachers College, Columbia University, 1925.

RUGG, H. O., "Statistical Methods Applied to Education," Houghton Mifflin Company, 1917.

SARGENT, D. A., "Physical Test of a Man," *School and Society,* January 29, 1921.

SCHWEGLER and ENGELHARDT, "A Test of Physical Efficiency," *American Physical Education Review,* November, 1924.

THORNDIKE, E. L., "Psychological Examination for High School Graduates," Teachers College, Columbia University, New York.

VINCENT, LEONA, "A Study of Intelligence Test Elements," Teachers College, Columbia University, New York, 1924.

WARDLAW, C. D., "Fundamentals of Baseball," Chapter on Testing the Mechanics of Baseball," by D. K. Brace, *Scribner's,* 1924.

WHIPPLE, G. M., "Manual of Mental and Physical Tests," Warwick and York, Baltimore, 1914.

WILLIAMS, J. F., "The Organization and Administration of Physical Education, Macmillan, 1920.

INDEX

Accomplishment, quotient, 89; formula for, 89, 90
Activities, Types sampled by tests, 6; relative value, 15
Achievement, tests in basket ball, 74; in indoor baseball, 78
Administering tests, 10
Age distributions: boys and girls, 129; college women, 130
Age norms, 9; for college women, 69; correlated with motor ability, 86, 87
Agility, 6; and balance, 6
Agreement, per cent of, 34
American, Physical Education Review, 13
Applications of scale of motor ability tests, 73; suggested, 63, 89; summary on, 98, 99

Bailor, E. M., 59
Balance tests, 6; in Forms C and D, 64
Baldwin, Long Island, 7, 17, 29
Baseball, indoor, achievement tests in, 78; correlated with intelligence, 89
Basket ball, achievement tests in, 74
Batting, 78
Booklet, test, 100
Brace, D. K., 67, 74, 78; Scale of Motor Ability Tests, 100

Classification, on basis of motor ability, 91
Conclusion: validity by judgment ratings, 20ff; validity by athletic events, 23; validity by scores of team members, 24; reliability of pupils scores, 33ff; mean net difference, Form A with B, 36; Correlation between first and second trials, 41; objectivity by self-scoring, 45; weighting on mean criterion scores, 52, 53; on elimination of tests, 62; on construction of equivalent forms, 66; correlation motor ability with baseball, 81; motor ability and soccer tests, 83; motor ability and track events, 86; motor ability and age, 87; motor ability and weight, 88; correlation motor ability with basket ball tests, 77; motor ability and intelligence, 89
Constancy, per cent of, 29, 30
Control, 6
Converting test scores into scale scores, ages 8-18, 126; college women, 127
Correlation, coefficients of: basket ball tests with intelligence, 16; tests passed with judgment ratings, 18, 25; Forms A and B with judgment ratings, 19;

table of, 20; motor ability tests with athletic events, 21ff; between equivalent forms, 35; self, 35; one year with next year, 38; failures on first trials with failures on second trials, 40; college women first trial with retrial, 41; criterion with tests weighted by validity, 57; depending on number of tests, 61; Form M with Form N, 65; motor ability with game skills, 73; motor ability with baseball, 80; motor ability with soccer tests, 81, 83; motor ability with track and field events, 85; motor ability with age, 86, 87; motor ability with weight, 87; motor ability with intelligence, 88
Criteria, for finding tests, 1; of motor ability, 13; judgment of motor ability ranked, 14; selection of, 16; judgment ratings, 17ff; of validity, 26; summary on, 94
Criterion, scores, 49; mean criterion scores, 51, 53; method of obtaining sum of, 56
Cubberly, H, 81

Decathlon tests, 22
Demonstration, of tests, 28
Description, of tests, 28; of scale tests, 105ff
Difficulty, tests weighted in accordance with, 50
Directions, general for applying tests, 100; for handling pupils, 100; when tests scored by examiner, 100ff; when pupils score themselves, 101; when pupils score each other, 101
Discobolus, 81
Distributions, age: boys and girls, 129; college women, 130
Dribble, and shoot, 74; dribbling soccer ball, 81

Engelhardt, 1
Equipment for tests, 1, 9; and reliability, 27
Equivalent forms, construction of, 63; groups, 91
Experimental, studied and motor ability, 91
Extending the range of T scale, 70

Failures: first trial with retrial, 29ff; on successive trials, 30
Flexibility, 6
Form A, 2ff; B, 4ff, 19; R and S, 53; C and D, 64; E and F, 65; M and N, 60, 97, 103; M, 105ff; N, 115ff
Formula, Pearson product-moment correlation, 15; probable error, 16; correlation ratio, 18; predictive index, 20, 59; Blakeman's test, 19; standard deviation, 37; McCall-Long method, 57
Frost, H., 81
Goal, soccer, 82; kicking, 82
Gucker, Colba, 22

Heading, soccer ball, 82
Hetherington, C. W., 1
Hillas, M., 77

Holzinger, K. J., 16
Horace Mann School, 7

Illustrations of Brace scale of motor ability tests, 105ff
Instructions for administering tests, 28, 93; on scoring, 105ff
Intelligence, correlated with motor ability, 88

Judgment rating, how given, 18, 94
Jump and reach, 75

Kelley, T. L., 18, 49, 54, 59
Kicking, soccer goals, 82

Lincoln School, 8, 44

McCall, W. A., 16, 29, 31, 32, 33, 35, 42, 49, 54; McCall-Long method, 54; steps in, 55, 63, 66, 67, 97
Mean, score of team members, 24; net difference, 29; net difference between Forms A and B, 36, 37; net difference over intervening year, 36, 38; tests passed, 129, 130
Meaning, of scores, 104
Motor Ability Test, 7
Motor accomplishment quotient, 89, 90

Net difference, 31, 33, 36ff
New Jersey College for Women, 8

Objectivity, determination of, 27, 42; definition of, 43; as relates to group scoring, 44; summary on, 95ff
Optimum scoring key, 48

Pass, push, 75; speed, 75
Percentile scales, 66
Pitch, underarm, 79
Practice, trials, 12
Predictive index, 59
Probable error, 16
Proceedings, Athletic Research Society, 13, 66
Psychological examination, 88

Reduction, of tests, 60, 62
Reliability, determination of, 27; shown by failures on retrial, 29; courses of unreliability, 31; by mean net difference, 32; of pupils scores, 33; summary on, 85ff
Retrial, failures on, 29
Rogers, F. R., 1, 87, 91
Rugg, H. O., 15, 16, 19, 21, 84

Sargent, D. A., 1, 75
Scale score, 104
Scaling the tests, 48, 97
Scattergram, 18, 19, 41, 131, 132
Schwegler, 1
Scores, on athletic events, 94; of team members, 94; as tests passed, 97; scale scores, 126, 127
Scoring, 18; key, 48; instructions, 105ff
Scoring blank, 125
Skidmore College, 8
Soccer, achievement tests in, 81
Shooting baskets, 74

Spearman, formula, 35
Speed pass, 75; throw, 78
Standard deviation, 37; 48
Stockdale, Carl, 18
Strength, 6; and control, 6
Subjects, the, 1, 7, 8, 94
Summary, Chapter I, 8; on validity, 25; Reliability and Objectivity, 45; on scaling the tests, 70; general, 93

T Scale, 67; construction of for 11-12-13-year-olds, 68; steps in construction of, 69; of basket ball tests, 76; baseball achievement tests, 80; soccer achievement tests, 84
Target, basket ball, 74; baseball, 79
Teachers College, subjects, 8, 39, 44
Techniques, in scaling tests, 96ff
Test Booklet, 71, 100
Tests, the, 1; descripition of tests tried, 2ff; nature of, 6; administering, 10; instructions, 10, 11, 100; demonstration, 11, 93; Test 1, 105; Test 2, 106; Test 3, 107; Test 4, 108; Test 5, 108ff; Test 6, 110; Test 7, 111; Test 8, 112; Test 9, 113; Test 10, 114; Test 11, 115; Test 12, 116; Test 13, 117; Test 14, 118; Test 15, 119; Test 16, 120; Test 17, 121; Test 18, 122; Test 19, 123; Test 20, 124.
Thorndike, E. L., 88
Throw, single overhand at target, 74; speed throw, 78; overhand, 79; throw in, 83

Validity of tests, 13, 18, 19, 20, 21, 22, 23, 24; shown by judgment rating, 17; shown by athletic events, 21; shown by scores of team members, 24; as a basis of weighting scores, 54; McCall-Long method, 54, 55, 57, 58; summary on, 94
Vincent, E. Leona, 54, 60

Wardlaw, C. D., 78
Weighting, in accordance with difficulty, 49; on basis of mean criterion scores, 49; on difference between mean criterion scores of success and failure, 52; on basis of validity, 54.
Whipple, G. M., 1
Williams, J. F., 1, 13

Zone system, 85

GV436 .B7
Measuring motor ability : a scale o
Gutman Library ANR4387
3 2044 028 700 292

DATE DUE

DEC 14 1979
FEB 22 1985

GAYLORD PRINTED IN U.S.A.

WITHDRAWAL

P8-BXM-001